09/508710

PCT/GB 98/02802

G

The Patent Office

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP9 1RH

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 29 September 1998

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

17SEP97 E303486-1 D00192
P01/7700 25.00 - 9719727.1

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form))

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

THE PATENT OFFICE
A
16 SEP 1997
LONDON

1. Your reference — N.73312 GCW.AB

2. Patent application number *(The Patent Office will fill in this part)* — 16 SEP 1997 — 9719727.1

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

RHONE-POULENC AGRICULTURE LIMITED
Fyfield Road
Ongar
Essex
CM5 0HW

Patents ADP number *(if you know it)* 5683206001

If the applicant is a corporate body, give the country/state of its incorporation — UNITED KINGDOM

4. Title of the invention — NEW PLANT GENES

5. Name of your agent *(if you have one)*

J A KEMP & CO

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

14 SOUTH SQUARE
GRAY'S INN
LONDON WC1R 5LX

26001

Patents ADP number *(if you know it)*

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*
-	-	-

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*
-	-

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer "Yes" if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body:
See note (d))

Yes

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	-
Description	86
Claim(s)	10
Abstract	1
Drawing(s)	2 + 2.

10. If you are also filing any of the following, state how many against each item.

Priority documents	-
Translations of priority documents	-
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	-
Request for preliminary examination and search *(Patents Form 9/77)*	1
Request for substantive examination *(Patents Form 10/77)*	-
Any other documents *(please specify)*	-

11. I/We request the grant of a patent on the basis of this application

Signature J.A. Kemp & Co. Date 16 September 1997

12. Name and daytime telephone number of person to contact in the United Kingdom

G.C. WOODS
0171 405 3292

NEW PLANT GENES

FIELD OF THE INVENTION

This invention relates to glutathione transferase (GST) subunits, to nucleic acid sequences encoding glutathione transferase subunits, and to uses of these glutathione transferases and coding sequences, especially in the field of plant biotechnology.

BACKGROUND OF THE INVENTION

Glutathione transferases (GSTs, EC. 2.5.1.18), also referred to as glutathione *S*-transferases, are multifunctional enzymes capable of catalysing the conjugation of electrophilic substrates with the tripeptide glutathione (GSH, gamma-glutamylcysteinylglycine). The electrophilic substrate may be of natural or synthetic origin, examples including endogenous stress-metabolites, drugs, pesticides and pollutants. Conjugation with GSH renders the compounds non-toxic and suitable for export from the cytosol and further metabolism. In addition to their activities in GSH conjugation, GSTs may have additional activities as glutathione peroxidases, catalysing the reduction of organic hydroperoxides to the corresponding alcohol according to the reaction:

R-OOH + 2 GSH -----> R-OH + GSSG.

All known active GSTs are composed of two polypeptide subunits, with each subunit possessing a binding site for GSH and the electrophilic co-substrate. The two subunits may either be identical, giving rise to a homodimer, or dissimilar giving rise to heterodimers.

GSTs may therefore be defined according to their source, or class, and their component subunits according to the nomenclature SpGST x-y, where Sp = source or class of GST; x and y describe the subunit types.

Each discrete subunit is encoded by a distinct gene, with many eucaryote containing GST multigene families encoding multiple isoenzymes.

The plant in which GSTs have been characterised in the greatest detail is maize (*Zea mays* L.). The major maize GSTs are composed of three discrete subunits, termed I, II and III. These subunits associate together to form three isoenzymes containing the *Zea mays* GST I subunit, namely *Zm*GSTI-I, *Zm*GSTI-II and *Zm*GSTI-III as well as the homodimers *Zm*GSTII-II and *Zm*GSTIII-III. The nucleotide sequences of *Zm*GSTI, *Zm*GSTII and *Zm*GSTIII have been determined. In view of their relatedness in sequence, these maize GSTs have collectively been termed type 1 plant GSTs. Additional maize GSTs with activities toward herbicides have been described as *Zm*GSTV-V and *Zm*GSTV-VI. The sequence of *Zm*GSTV differs markedly from the other maize GSTs described to date, resembling the auxin-inducible GSTs from dicotyledenous plants which have been termed the type 3 GSTs.

The maize GST subunit types are associated with differing substrate specificities. The *Zm*GSTI subunit has broad-ranging, but low, activities toward chloro-*s*-triazine, chloroacetanilide and diphenyl ether herbicides. The *Zm*GSTII and *Zm*GSTIII subunits show greater specificity toward chloroacetanilides, while *Zm*GSTV and *Zm*GSTVI are highly active toward diphenyl ethers. The GST isoenzymes differ in their patterns of expression in the organs of maize. Thus, *Zm*GSTI-I and

*Zm*GSTV-V are expressed in all plant parts, while *Zm*GSTI-II is root specific. The expression of the GST subunits is also differentially affected by herbicide safeners. These are compounds which enhance the tolerance of cereal crops to herbicides, in part, by increasing the expression of detoxifying enzymes such as GSTs. Thus, the *Zm*GSTII and *Zm*GSTV subunits accumulate in maize seedlings following treatment with the safeners dichlormid or benoxacor while the *Zm*GSTI and *Zm*GSTIII subunits are only modestly enhanced by safeners.

Far less is known regarding GSTs in plant species other than maize. GSTs with activities toward non-herbicide substrates have been identified in some plants, and mRNAs apparently encoding GSTs have been shown to be expressed in plants including carnation, tobacco and thale cress (*Arabidopsis thaliana*). However, isoenzymes with activities toward herbicides have only been definitively identified in soybean, pea and pine trees. Of these, only in soybean has the nucleotide coding sequences of the herbicide-detoxifying GST been reported.

GSTs in plants have also been shown to have secondary activities as glutathione peroxidases, able to reduce organic hydroperoxides, such as fatty acid hydroperoxides to the corresponding monohydroxy alcohols. GSTs with glutathione peroxidase activity have been isolated from peas, soybean, *A. thaliana* and wheat flour. Since fatty acid hydroperoxides are a common result of membrane peroxidation imposed during oxidative stress, glutathione peroxidases provide an important cytoprotective function in preventing the accumulation of fatty acid hydroperoxides and their subsequent degradation to toxic aldehydes. Glutathione peroxidases may therefore have a vital function in protecting plant

cells from oxidative stress. The intervention of glutathione peroxidases in lipid peroxidation has also been cited as a determinant of flour quality in wheat.

Of particular relevance to this invention is the lack of knowledge concerning the GSTs of wheat (*Triticum aestivum* L.).

Some information is available from experiments on whole plants and plant extracts. Several herbicides including examples of the chloroacetanilides, as well as dimethenamid and fenoxaprop-ethyl undergo GSH conjugation in the course of their detoxification in wheat. Also, in crude plant extracts GST activities toward chloroacetanilide herbicides and fenoxaprop-ethyl have been demonstrated.

However, there is only one previous report of the purification of a GST from wheat. This GST, purified from wheat flour, was described as a homodimer of 27.5 kDa polypeptides with activity toward the non-herbicide substrate 1-chloro-2,4-dinitrobenzene (CDNB) and glutathione peroxidase activity toward fatty acid hydroperoxides.

Moreover, very little is known regarding GST genes in wheat. An mRNA originally described as *wir5*, which showed sequence similarity to the type 1 maize GSTs, was identified as accumulating in wheat leaves during the onset of acquired resistance to powdery mildew (*Erysiphe graminis)*. The gene was termed *gstA1* and shown to be similar in genomic organisation to maize *ZmGST1*. The gstA1 polypeptide was expressed in recombinant bacteria and shown to have an apparent molecular mass of 29 kDa. The respective enzyme showed GST activity towards the

non-herbicide CDNB, though the activity toward other substrates and activity as a glutathione peroxidase was not reported. An antibody was raised to the recombinant GstA1 and used in Western blotting experiments to show that this GST was specifically induced in wheat leaves by pathogen attack. In contrast, a distinct class of GSTs composed of 25 kDa and 26kDa subunits, which were recognised by an antiserum raised to undefined GSTs in maize, accumulated following exposure to cadmium and the herbicides atrazine, alachlor and paraquat. The activities of these xenobiotic-inducible GSTs in wheat and the corresponding nucleotide sequences were not reported.

Thus, although wheat is an important crop plant, there has been almost no molecular characterisation of wheat GSTs or their genes and, to date, only one purified GST and one GST gene sequence is available, *gst*A1.

Significantly, the purified recombinant GST protein expressed from gene *gstA1* was not reported to exhibit activity towards herbicides. Hence, none of the previous work on wheat GSTs actually provides any means of achieving herbicide resistance based on the function of wheat GSTs.

SUMMARY OF THE INVENTION

We have purified four GST isoenzymes with activity toward herbicides from wheat shoots treated with the herbicide safener fenchlorazole-ethyl and have identified four distinct subunits. In safener-treated shoots, we have found that the predominant GST subunit is a 25 kDa polypeptide, which has been termed *Triticum aestivum* GST 1 (*Ta*GST1). Additionally, two distinct 26 kDa subunits

have been identified and termed *Ta*GST2 and *Ta*GST3 and a 24 kDa subunit, termed *Ta*GST4. These subunits associate together to form the active dimeric isoenzymes *Ta*GST1-1, *Ta*GST1-2, *Ta*GST1-3 and *Ta*GST1-4.

In our experiments, the expression of all four isoenzymes was affected by the herbicide safener fenchlorazole-ethyl, although each one responds in a somewhat different way. The *Ta*GST1-1 isoenzyme is the major GST present in the leaves of untreated wheat seedlings, and its expression is increased by approximately 50% following exposure to fenchlorazole-ethyl. *Ta*GST1-4 is expressed at low levels in untreated shoots and its expression is greatly increased by safener application, while *Ta*GST1-2 and *Ta*GST1-3 are only observed following treatment with the safener. All four of these GST isoenzymes have broad-ranging activities toward xenobiotic substrates and all four demonstrate activity towards herbicides and additional activities as glutathione peroxidases able to reduce organic hydroperoxides, with *Ta*GST1-4 being the most active in this respect. Each isoenzyme also has specific properties. Thus, for example, detoxification of one particular herbicide, fenoxaprop-ethyl, is associated with the more strongly safener-inducible *Ta*GST1-2, *Ta*GST1-3 and *Ta*GST1-4 heterodimers, rather than with the *Ta*GST1-1 homodimer.

Furthermore, we have identified, cloned and sequenced cDNAs for the major type 3 GSTs in wheat, together with cDNAs encoding a range of type I GSTs, all active in herbicide metabolism. This is fundamental to understanding the GST detoxification system in wheat and to exploiting it to generate transgenic herbicide-resistant plants expressing wheat GSTs. In many previous

studies, GST activity could not be linked to specific genes, precluding this approach.

From the sequences of the cDNAs the amino acid sequences of the GST subunits themselves has been deduced.

Accordingly, the invention provides:

a polynucleotide encoding a glutathione transferase (GST) subunit, which polynucleotide comprises a coding sequence capable of hybridising selectively to the coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 or to the complement of one of those sequences.

The invention also provides:

a polypeptide which is a GST subunit and comprises the amino acid sequence of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14 or a sequence substantially homologous thereto, or a fragment of either said sequence.

The invention also provides:

a dimeric protein comprising two GST subunits, wherein at least one subunit is a polypeptide of the invention.

The invention also provides:

a chimeric gene comprising a polynucleotide of the invention operably linked to regulatory sequences that allow expression of the coding sequence in a host cell.

The invention also provides:

a vector comprising a polynucleotide of the invention or a chimeric gene of the invention.

The invention also provides:

a cell transformed or transfected with a vector of the invention.

The invention also provides:

a cell having, integrated into its genome, a chimeric gene of the invention.

The invention also provides:

a process for the production of a polypeptide of the invention, which process comprises:

(a) cultivating a cell of the invention under conditions that allow the expression of the polypeptide; and

(b) recovering the expressed polypeptide.

The invention also provides:

a process for the production of a dimeric protein of the invention, which process comprises:

(a) cultivating a cell of the invention under conditions that allow:

(i) the expression of the polypeptide of the invention and, if a further polynucleotide sequence as defined herein is present, optionally the expression of a

further GST subunit encoded by a further polynucleotide, and

(ii) the association of the GST subunit polypeptide of the invention with another GST subunit polypeptide to form a dimeric protein of the invention; and

(b) recovering the dimeric protein so formed.

The invention also provides:

a method of obtaining a transgenic plant cell comprising:

(a) transforming a plant cell with an expression vector of the invention to give a transgenic plant cell,

and optionally,

(a') transforming the cell with one or more further polynucleotide sequences coding for a GST subunit, operably linked to regulatory elements that allow expression of the subunit in the cell.

The invention also provides:

a method of obtaining a first-generation transgenic plant comprising:

(b) regenerating a transgenic plant cell transformed with a vector of the invention to give a transgenic plant.

The invention also provides:

a method of obtaining a transgenic plant seed

comprising:

(c) obtaining a transgenic seed from a transgenic plant obtainable by regenerating a transgenic plant cell transformed with a vector of the invention.

The invention also provides:

a method of obtaining a transgenic progeny plant comprising obtaining a second-generation transgenic progeny plant from a first-generation transgenic plant obtainable by regenerating a transgenic plant cell transformed with a vector of the invention, and optionally obtaining transgenic plants of one or more further generations from the second-generation progeny plant thus obtained.

The invention also provides:

a method of obtaining a transgenic progeny plant comprising obtaining a second-generation transgenic progeny plant from a first-generation transgenic plant obtainable by regenerating a transgenic plant cell transformed with a vector of the invention comprising:

(c) obtaining a transgenic seed from a first-generation transgenic plant obtainable by regenerating a transgenic plant cell transformed with a vector of the invention, then obtaining a second-generation transgenic progeny plant from the transgenic seed;

and/or

(d) propagating clonally a first-generation transgenic plant obtainable by regenerating a transgenic plant cell

transformed with a vector of the invention to give a second-generation progeny plant;

and/or

(e) crossing a first-generation transgenic plant obtainable by regenerating a transgenic plant cell transformed with a vector of the invention with another plant to give a second-generation progeny plant;

and optionally;

(f) obtaining transgenic progeny plants of one or more further generations from the second-generation progeny plant thus obtained.

The invention also provides:

a transgenic plant cell, first-generation plant, plant seed or progeny plant obtainable by a method of the invention.

The invention also provides:

a transgenic plant or plant seed comprising plant cells of the invention.

The invention also provides:

a transgenic plant cell callus comprising plant cells of the invention, or obtainable from a transgenic plant cell, first-generation plant, plant seed or progeny plant of the invention.

The invention also provides:

use of a polynucleotide of the invention as a selectable marker for detecting transformation of a plant cell.

The invention also provides:

a nucleic acid construct comprising:

(a) a polynucleotide of the invention operably linked to regulatory elements that allow expression of the coding sequence in a plant cell; and

(b) a site into which a further polynucleotide comprising a coding sequence can be inserted.

The invention also provides:

a vector comprising such a construct.

The invention also provides:

a method of transforming a plant cell or of obtaining a plant cell culture or transgenic plant comprising:

(a) providing an untransformed plant cell which is susceptible to a herbicide whose herbicidal activity is reduced by a dimeric protein of the invention;

(b) transforming the plant cell with a vector comprising:

(i) a polynucleotide of the invention operably linked to regulatory elements that allow expression of the coding sequence in a plant cell; and

(ii) a site into which a further polynucleotide comprising a coding sequence can be inserted;

(c) cultivating the transformed cell under conditions that allow the expression of the polynucleotide (a) in the construct; and/or

(c') regenerating the cell to give a cell culture or plant such that the polynucleotide (a) in the construct is expressed; and

(d) contacting the cell, cell culture or plant with the herbicide whose herbicidal activity is reduced by the dimeric protein of the invention, and to which the untransformed plant cell was susceptible; and

(e) selecting cells, cell cultures or plants that are less susceptible to the herbicide than are corresponding untransformed cells, cell cultures or plants.

The invention also provides:

use of a dimeric protein of the invention in a method of identifying compounds capable of metabolism by a GST.

The invention also provides:

a method of identifying compounds capable of being metabolised by a glutathione transferase comprising:

(a) contacting a candidate compound suspected of being capable of being metabolised by glutathione transferase with glutathione (GSH) in the presence of a dimeric

protein of the invention; and

(b) determining whether or not metabolism of the candidate compound takes place.

The invention also provides:

compounds identified by such methods.

The invention also provides:

a kit for detecting compounds capable of being metabolised by a GST comprising:

(a) reduced glutathione or homoglutathione; and

(b) a dimeric protein of the invention.

The invention also provides:

an antibody which specifically recognises a polypeptide or dimeric protein of the invention.

The invention also provides:

a nucleic acid probe which selectively hybridises to the sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13.

The invention also provides:

a method of identifying compounds that induce GST expression in graminaceous plants comprising:

(a) contacting a graminaceous plant, or a cell or cell

culture thereof, with a candidate compound suspected of being capable of inducing GST expression; and

(b) determining the level of GST expression in the plant, cell or cell culture.

The invention also provides:

compounds identified by such methods.

The invention also provides:

a kit for identifying compounds that induce GST expression in plants by such a method, which kit comprises an antibody of the invention.

The invention also provides:

a method of determining the GST level in a sample of seed or flour comprising:

(i) determining the level of GST protein present by using an antibody of the invention; or

(ii) determining the level of GST mRNA present using a probe of the invention.

The invention also provides:

a method of controlling the growth of weeds at a locus where a transgenic plant of the invention is being cultivated, which method comprises applying to the locus a herbicide whose herbicidal properties are reduced by a dimeric protein of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Anion-exchange chromatography of affinity-purified wheat GSTs.

Chromatography of A: affinity-purified polar GSTs; and B: affinity-purified hydrophobic GSTs on Hi-Trap Q-Sepharose columns eluted with the increasing NaCl gradient shown. The eluent was monitored for A_{280} as shown with the unbroken line and individual fractions assayed for GST activity.

Figure 2. HPLC analysis of wheat GST subunits.

Reversed-phase HPLC analysis of polypeptide subunits present in A, affinity-purified polar GSTs; B, affinity-purified hydrophobic GSTs; C, the isoenzyme *Ta*GST1-1, resolved by anion-exchange chromatography of the affinity-purified polar GSTs.

DETAILED DESCRIPTION OF THE INVENTION

Polynucleotides

The invention provides polynucleotides comprising sequences encoding novel GST subunits, SEQ ID Nos 1, 3, 5, 7, 9, 11 and 13 and sequences that hybridise selectively to these coding sequences thereof or their complementary sequences. It also provides polynucleotide fragments of these sequences that encode polypeptides having GST activity, as defined herein.

A polynucleotide of the invention is capable of hybridising selectively with the coding sequence of SEQ

ID No. 1, 3, 5, 7, 9, 11 or 13 or to the sequence complementary to one of those coding sequences. Polynucleotides of the invention include variants of the coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 which can function as GSTs, when dimerised with another GST subunit. Typically, a polynucleotide of the invention is a contiguous sequence of nucleotides which is capable of selectively hybridising to the coding sequence of SEQ ID. No. 1, 3, 5, 7, 9, 11 or 13 or to the complement of that coding sequence.

A polynucleotide of the invention can hybridise to coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 at a level significantly above background. Background hybridisation may occur, for example, because of other cDNAs present in a cDNA library. The signal level generated by the interaction between a polynucleotide of the invention and the coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 is typically at least 10 fold, preferably at least 100 fold, as intense as interactions between other polynucleotides and the coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P. Selective hybridisation is typically achieved using conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

A nucleotide sequence capable of selectively hybridising to the DNA coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 or to the sequence complementary to one of those coding sequences will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95%, 98% or 99%, homologous to the coding sequence

of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 or the complement of one of those sequences over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 90% homologous over 25, preferably over 30 nucleotides forms one aspect of the invention, as does a polynucleotide which is at least 95% homologous over 40 nucleotides.

Polynucleotides of the invention may comprise DNA or RNA. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to polynucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or lifespan of polynucleotides of the invention.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers,

probes and other fragments will preferably be at least 10, preferably at least 15 or 20, for example at least 25, 30 or 40 nucleotides in length.

Polynucleotides such as a DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques. The polynucleotides are typically provided in isolated and/or purified form.

In general, primers will be produced by synthetic means, involving a stepwise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Genomic clones corresponding to the cDNAs of SEQ ID No. 1, 3, 5, 7, 9, 11 and 13 containing, for example introns and promoter regions are also aspects of the invention and may also be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques, starting with genomic DNA from a wheat (*Triticum aestivum L.)*, cell, e.g. a wheat shoot cell or a cell of a plant of a related *Triticum* species, for example as described by Feldman *et al.*, (Scientific American, (1981), vol. 244(1) pages 98 to 109).

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al*, 1989, Molecular Cloning: a laboratory manual.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the

scope of the invention can be obtained in a number of ways.

Other allelic variants of the wheat sequences of SEQ ID Nos. 1, 3, 5, 7, 9, 11 and 13 including those from *Triticum aestivum L.* species itself related to *Triticum aestivum L.* (cf Feldman *et al, supra*) may be obtained for example by probing genomic DNA libraries made from a range of wheat cells, using probes as described above.

In addition, other plant homologues of SEQ ID Nos. 1, 3, 5, 7, 9, 11 and 13 may be obtained and such homologues and fragments thereof in general will be capable of selectively hybridising to the coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 or its complement. Such sequences may be obtained by probing cDNA or genomic libraries from other plant species with probes as described above. Degenerate probes can be prepared by means known in the art to take into account the possibility of degenerate variation between the DNA sequences of SEQ ID Nos. 1, 3, 5, 7, 9, 11 and 13 and the sequences being probed for under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

Allelic variants and species homologues may also be obtained using degenerate PCR which will use primers designed to target sequences within the variants and homologues encoding likely conserved amino acid sequences. Likely conserved sequences can be predicted from aligning the amino acid sequences of the invention (SEQ ID No. 2, 4, 6, 8, 10, 12 and 14) with that of other similar GST subunit sequences. The primers will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning

sequences with single sequence primers against known sequences.

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of SEQ ID Nos. 1, 3, 5, 7, 9, 11 or 13 sequences or allelic variants thereof. This may be useful where, for example, silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides.

The invention further provides double stranded polynucleotides comprising a polynucleotide of the invention and its complement.

Polynucleotides, probes or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, enzyme labels, or other protein labels such as biotin. Such labels may be added to polynucleotides, probes or primers of the invention and may be detected using techniques known *per se*.

The present invention also provides polynucleotides encoding the polypeptides of the invention described below. Because such polynucleotides will be useful as sequences for recombinant production of polypeptides of the invention, it is not necessary for them to be selectively hybridisable to the coding sequence of sequence SEQ ID Nos. 1, 3, 5, 7, 9, 11 or 13 although this will generally be desirable. Otherwise, such polynucleotides may be labelled, used, and made as

described above if desired. Polypeptides of the invention are described below.

Particularly preferred polynucleotides of the invention are those of SEQ ID No. 1, 3, 5, 7, 9, 11 and 13 and the polynucleotides that are the coding regions within those sequences i.e. the regions which encode the polypeptides of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14.

Polypeptides

A polypeptide of the invention consists essentially of the amino acid sequence set out in SEQ ID No. 2, 4, 6, 8, 10, 12 or 14 or a substantially homologous sequence, or of a fragment of either of these sequences. In general, the naturally occurring amino acid sequences shown in SEQ ID Nos. 2, 4, 6, 8, 10, 12 and 14 are preferred. However, the polypeptides of the invention include homologues of the natural sequences, and fragments of the natural sequences and of their homologues, which have GST activity.

The polypeptides of the invention are glutathione transferase (GST) subunits. The invention also provides dimeric proteins comprising two GST subunits wherein at least one subunit is a polypeptide of the invention.

Thus, the polypeptides of the invention are normally functionally active as GSTs when dimerised with another GST subunit. Thus, dimeric proteins of the invention are capable of catalysing the conjugation of the tripeptide glutathione (GSH, gamma-glutamylcysteinyl glycine) and/or related derivatives to an electrophilic substrate of natural or synthetic origin. Related derivatives include homoglutathione (gamma-glutamylcysteinyl alanine) and

hydroxyglutathione (gamma-glutamylcysteinyl serine).

Optionally, they may also have one or more of the other properties of naturally occurring GSTs including glutathione peroxidase activity as defined above.

Preferably, they have GST activity towards one or more herbicide substrates. For example, they may have activity towards one or more of the following herbicides: Fluorodifen, Fenoxaprop-ethyl, Metolachlor, Alpha-Metolachlor, Acetochlor, Alachlor, Pretilachlor, Fluthiamid, Dimethenamid, S-Dimethenamid, Flupyrsulfuron-methyl, Triflusulfuron-methyl, Acifluorfen, Chlorimuron-ethyl, Fomesafen, Atrazine, Simazine, Cyanazine, Metribuzin. Particularly preferred herbicides include Fenoxaprop-ethyl, Flupyrsulfuron-methyl, Fluthiamid, Acetochlor, Metolachlor and Alpha-Metolachlor.

Most preferably, a dimeric protein of the invention is able to catalyse the conjugation of GSH to one or more of the following herbicide substrates: Fenoxaprop-ethyl, Flupyrsulfuron-methyl, fluthiamid, Acetochlor, Metolachlor and Alpha-Metolachlor.

Optionally, a dimeric protein of the invention may be able to catalyse the conjugation of GSH to one or more non-herbicide substrates, for example CDNB. They may also have activity towards phytotoxic non-herbicide substrates.

Optionally, monomeric polypeptides of the invention may have GST activity as defined above, even when not dimerised.

In particular, a polypeptide of the invention may

comprise:

(a) the polypeptide sequence of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14;

(b) an allelic variant or species homologue thereof; or

(c) a protein at least 70 80, 90, 95, 98 or 99% homologous to (a) or (b).

An allelic variant will be a variant which will occur naturally in a plant and which will function in a substantially similar manner to the protein of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14, as defined above. Similarly, a species homologue of the protein will be the equivalent protein which occurs naturally in another plant species which can function as GST. Such a homologue may occur in plants other than wheat, particularly monocotyledonous plants such as related *Triticum* species, rice, maize, oats, rye, barley, triticale or sorghum. Within any one species, a homologue may exist as several allelic variants, and these will all be considered homologues of the protein of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14.

Allelic variants and species homologues can be obtained by following the procedures described herein for the production of the polypeptides of SEQ ID No. 2, 4, 6, 8, 10, 2 and 14 and performing such procedures on a suitable cell source e.g. a cell of a wheat genotype carrying an allelic variant, or a cell of a plant of a different another species. It will also be possible to use a probe as defined above nucleotide sequence to probe libraries made from plant cells in order to obtain clones encoding the allelic or species variants. The clones can be manipulated by conventional techniques to generate a polypeptide of the invention which can then be produced

by recombinant or synthetic techniques known *per se*.

A polypeptide of the invention is preferably at least 70% homologous to the protein of SEQ ID No. 2, 4, 6, 8, 10, 2 or 14, more preferably at least 80 or 90% and more preferably still at least 95%, 97% or 99% homologous thereto over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous amino acids. Methods of measuring protein homology are well known in the art and it will be understood by those of skill in the art that in the present context, homology is calculated on the basis of amino acid identity (sometimes referred to as "hard homology").

The sequence of the polypeptides of SEQ ID Nos 2, 4, 6, 8, 10, 12 and 14 and of allelic variants and species homologues can thus be modified to provide polypeptides of the invention.

Amino acid substitutions may be made, for example from 1, 2 or 3 to 10, 20 or 30 substitutions. The modified polypeptide generally retains activity as a GST, as defined herein. Conservative substitutions may be made, for example according to the following Table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other.

ALIPHATIC	Non-polar	G A P
		I L V
	Polar-uncharged	C S T M
		N Q
	Polar-charged	D E
		K R
AROMATIC		H F W Y

Polypeptides of the invention also include fragments of the above-mentioned full length polypeptides and variants thereof, including fragments of the sequence set out in SEQ ID No. 2, 4, 6, 8, 10, 12 and 14. Such fragments typically retain activity as a GST.

Other preferred fragments include those which include an epitope. Suitable fragments will be at least about 5, e.g. 10, 12, 15 or 20 amino acids in size. Polypeptide fragments of the polypeptides of SEQ ID Nos. 2, 4, 6, 8, 10, 12 and 14, and allelic and species variants thereof may contain one or more (e.g. 2, 3, 5, or 10) substitutions, deletions or insertions, including conserved substitutions. Epitopes may be determined either by techniques such as peptide scanning techniques already known in the art. These fragments will be useful for obtaining antibodies to polypeptides and dimeric proteins of the invention.

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated. A polypeptide of the invention may also be in a

substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

Polypeptides of the invention may be modified for example by the addition of Histidine residues or a T7 tag to assist their identification or purification or by the addition of a signal sequence to promote their secretion from a cell.

A polypeptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I, ^{35}S, enzymes, antibodies, polynucleotides and linkers such as biotin.

Polypeptides and dimeric proteins of the invention may be chemically modified, e.g. post-translationally modified. For example, they may be glycosylated or comprise modified amino acid residues. Such modified polypeptides and proteins fall within the scope of the terms "polypeptide" and "dimeric protein" of the invention.

Dimeric proteins

The invention also provides dimeric proteins having two GST subunits wherein at least one of the two subunits is a polypeptide of the invention. These dimeric proteins may have two identical subunits of the invention, i.e. they may be homodimeric. Alternatively, they may have two dissimilar subunits; i.e. they may be heterodimeric.

In heterodimers, the two subunits may both be polypeptides of the invention. Alternatively, one subunit may be a polypeptide of the invention, whilst the other is a different GST subunit.

Thus, for example, heterodimeric proteins of the invention may have one subunit which is a polypeptide of the invention, and one which is a known GST subunit from maize (e.g. ZmGSTI, ZmGSTII, ZmGSTIII, ZmGSTIV, ZmGSTV or ZmGSTVI: see above), or another species.

Preferably, the dimeric proteins have two subunits that are polypeptides of the invention. Various combinations of polypeptides of the invention are possible. Preferred combinations include:

*Ta*GST1-1 (SEQ ID No. 2/SEQ ID No. 2);
*Ta*GST1-2 (SEQ ID No. 2/SEQ ID No. 4);
*Ta*GST1-3 (SEQ ID No. 2/SEQ ID No. 6);
*Ta*GST1-4 (SEQ ID No. 2/SEQ ID No. 8);

being representative of the major combinations found in GSTs in wheat.

The invention also provides dimeric proteins having two subunits as described above which are fusion proteins. In these fusion proteins, the two subunits are joined by a linker polypeptide. Any linker may be used as long as it does not interfere significantly with the correct association of the two subunits or with the GST activity of the dimer. Such fusion proteins will typically be prepared by joining together the polynucleotides encoding the two monomers in the correct reading frame, then expressing the composite

polynucleotide coding sequence under the control of regulatory sequences as defined herein. These composite polynucleotide coding sequences are a further aspect of the invention, as are chimeric genes and vectors comprising them, methods of producing them by recombinant means, and cells and plants comprising such vectors or chimeric genes. It will be understood that dimeric proteins of the invention may be such fusion proteins.

Vectors and chimeric genes

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and cultivating the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors. Bacterial cells, especially *E. Coli* are preferred.

Expression vectors

Preferably, a polynucleotide of the invention in a vector is operably linked to regulatory sequences capable of effecting the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. Such expression vectors can be used to express the polypeptides of the invention.

The term "operably linked" refers to a juxtaposition

wherein the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is positioned in such a way that expression of the coding sequence is achieved under conditions compatible with the regulatory sequences.

Such vectors may be introduced into a suitable host cell to provide for expression of a polypeptide or polypeptide fragment of the invention, as described below.

The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, preferably a promoter for the expression of the said polynucleotide and optionally an enhancer and/or a regulator of the promoter. For expression in plant cells, one preferred enhancer is the Tobacco etch virus (TEV) enhancer. A terminator sequence may also be present, as may a polyadenylation sequence. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene (e.g. nptI or nptII) or methotrexate resistance gene for a plant vector. Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell. The vector may also be adapted to be used *in vivo*, for example for generation of transgenic plants of the invention.

So far as plasmid vectors are concerned, plasmids derived from the Ti plasmid of *Agrobacterium tumefaciens* are especially preferred, as are plasmids derived from the Ri plasmid of *Agrobacterium rhizogenes*.

A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of polynucleotides of the invention. The cells will be chosen to be compatible with the said vector and may for example be prokaryotic (bacterial), plant, yeast, insect or mammalian cells, bacterial and plant cells being preferred.

Polynucleotides according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense polynucleotides may also be produced by synthetic means. Such antisense polynucleotides may be used in a method of controlling the levels of GSTs having the sequence of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14 or their variants or species homologues *in planta*.

Promoters and other regulatory elements may be selected to be compatible with the host cell for which the expression vector is designed.

Promoters suitable for use in plant cells may be derived, for example, from plants or from bacteria that associate with plants or from plant viruses. thus, promoters from *Agrobacterium spp*. including the nopaline synthase (nos), octopine synthase (ocs) and mannopine synthase (mas) promoters are preferred. Also preferred are plant promoters such as the ribulose bisphosphate small subunit promoter (rubisco ssu), and the phaseolin. promoter. Also preferred are plant viral promoters such as the cauliflower mosaic virus (CAMV) 35S and 19S promoters.

Depending on the pattern of expression desired, promoters may be constitutive or inducible. For example, strong constitutive expression in plants can be obtained with the CAMV 35S or rubisco ssu promoters. Also, tissue-specific or stage-specific promoters may be used to target expression of polypeptides of the invention to particular tissues in a transgenic plant or to particular stages in its development. Chemically inducible promoters such as those activated by herbicide safeners may also be used,for example the maize GST 27 promoter (WO97/11189), the maize In2-1 promoter (WO90/11361), the maize In2-2 promoter (De Veylder *et al*, Plant cell Physiology, Vol. 38, pp568-577 (1997).

Especially where expression in plant cells is desired, other regulatory signals may also be incorporated in the vector, for example a terminator and/or polyadenylation site. One preferred terminator is the nos terminator although other terminators functional is the nos terminator in plant cells may also be used.

Additionally, sequences encoding secretory signals or transit peptides may be included. On expression, these elements direct secretion from the cell or target the polypeptide of the invention to a particular location within the cell. For example, sequences may be added to target the expressed polypeptide to the nucleus or plastids (e.g. chloroplasts) of a plant cell.

Chimeric genes

The invention also provides chimeric genes suitable for securing the expression of polypeptides of the invention in a host cell, preferably a plant cell. These comprise a polynucleotide of the invention, operably

linked to regulatory sequences that allow its expression in a host cell, preferably a plant cell.

Typically, therefore, a chimeric gene comprises the following elements in 5' to 3' orientation: a promoter functional in a host (preferably plant) cell, as defined above, a polynucleotide of the invention and a terminator functional in said cell, as defined above. Other elements, for example an enhancer, may also be present. These chimeric genes may be incorporated into vectors, as defined above.

Expression in host cells

Expression vectors of the invention may be introduced into host cells using conventional techniques including calcium phosphate precipitation, DEAE-dextran transfection, or electroporation. For plant cells, preferred transformation techniques include electroporation of plant protoplasts, transformation by *Agrobacterium tumefaciens* and particle bombardment. Particle bombardment is particularly preferred for transformation of monocot cells.

Expression in the host cell may be transient although, preferably, integration of the polynucleotide or chimeric gene of the invention into the cell's genome is achieved.

Suitable cells include cells in which the above-mentioned vectors may be expressed. These include microbial cells such as bacteria such as *E. coli*, plant cells, mammalian cells such as CHO cells, COS7 cells or Hela cells, insect cells or yeast such as *Saccharomyces*.

Bacterial and plant cells are preferred.

Optionally, cells of the invention may comprise one or more further polynucleotide sequences encoding a GST subunit, operably linked to regulatory sequences, as defined above, that allow expression of the subunit in the cell. Such polynucleotide sequences may be further polynucleotides of the invention or they may encode other GST subunits as defined above with respect to dimeric proteins.

Such polynucleotides may be naturally present in the cell, e.g. if it is a plant cell or they may be introduced artificially, e.g. as defined above.

Such cells allow the production of heterodimeric proteins of the invention where the polynucleotides encode different GST subunits, or the production of monomeric polypeptides of the invention and/or homodimeric proteins of the invention in greater quantities. For example, they may allow the expression of active heterodimeric enzymes.

Cell culture will take place under standard conditions. Commercially available cultural media for cell culture are widely available and can be used in accordance with manufacturers' instructions.

Processes for production of polypeptides and dimeric proteins

The invention provides processes for the production of polypeptides and dimeric proteins of the invention by recombinant means.

Generally, monomeric GST subunits of the invention spontaneously dimerise to form homodimers and/or heterodimers of the invention. Thus, in general, expression of polypeptides of the invention gives rise to dimers in the first instance. These dimers may be the desired product; alternatively, it may be desirable to separate the monomers. For example, as described below, it may be desired to separate the monomeric subunits of a homodimer in order to combine them with different monomeric subunits, thereby yielding heterodimers.

Processes for the production of polypeptides of the invention may comprise:

(a) cultivating a transformed cell as defined above under conditions that allow the expression of the polypeptide;

and preferably

(b) recovering the expressed polypeptide.

For example, the expressed monomeric peptides may be recovered by denaturation of dimers formed by them, which separates the subunits. Then, the monomers can be recovered and renatured. Typically, they will then redimerise.

Processes for production of dimeric proteins of the invention may comprise:

(a) cultivating a transformed cell as defined above under conditions that allow

(i) the expression of the polypeptide of the invention and, if a further GST subunit-encoding sequence as defined above is present, optionally the expression of a further GST subunit encoded by the further sequence

and preferably

(ii) the association of the GST subunit polypeptide of the invention with another identical GST subunit polypeptide to form a home dimeric protein of the invention; and/or

(ii) the association of the GST subunit polypeptide of the invention with a non-identical GST subunit to form a heterodimeric protein of the invention.

and preferably

(b) recovering the dimeric proteins so formed, and optionally resolving them.

Where only a single type of GST subunit-encoding sequence of the invention is present in the transformed cell, these processes normally give rise to homodimeric proteins of the invention. Where one or more further GST subunit-encoding sequences is present, these processes give rise to heterodimers or to a mixture of some or all of the following: homodimers of each possible type.

Alternatively, dimeric proteins of the invention can be produced by expressing the required polypeptide subunits in separate cells. This typically leads to the production of two different types of homodimer. The desired heterodimer can then be prepared by: mixing the homodimers and denaturing the mixed sample, or by denaturing the homodimers separately and then mixing them; then renaturing the mixed sample. This will typically lead to a mixture of dimeric proteins comprising both possible types of homodimers and also heterodimers comprising one subunit of each type. Similarly, mixtures of greater numbers of types of dimer can be produced in this way if different homodimers are produced in three or more different cells, or if cells that give rise to heterodimers are used.

For these processes, any transformed cell as described above may be used. Bacterial cells are preferred, especially cells of *E. coli*, although other cell types may also be used.

Optionally, the polypeptide or dimeric protein may be isolated and/or purified, by techniques known in the art.

In processes of the invention, any suitable method may be used to denature and/or renature polypeptides of the invention, and suitable methods are well known in the art.

Similarly, where a mixture of polypeptide subunits or dimeric proteins results, these may be resolved or separated by any suitable technique known in the art.

Antibodies

The invention also provides monoclonal or polyclonal antibodies which specifically recognise polypeptides of the invention or dimeric proteins of the invention.

Thus, antibodies of the invention bind specifically to the polypeptides and/or dimers of the invention, preferably to the extent that they distinguish between the polypeptides and/or dimers of the invention and other GST subunits and GSTs.

Monoclonal antibodies may be prepared by conventional hybridoma technology using polypeptides or dimeric proteins of the invention as immunogens. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention and recovering immune serum. In order that such antibodies may be made, polypeptides may be haptenised to another polypeptide for use as immunogens in animals or humans. For the purposes of this invention, the term "antibody" includes antibody fragments such as Fv, F(ab) and $F(ab)_2$ fragments, as well as single chain antibodies.

Methods of producing transgenic plant cells, plant parts and tissues, plants and seeds of the invention

Transgenic plant cells, plant parts and tissues, plants and seeds of the invention are transgenic in the sense that the y have at least one polynucleotide of the invention introduced into them.

The invention provides a method of obtaining a transgenic plant cell comprising transforming a plant cell with an expression vector of the invention to give a transgenic plant cell; and optionally transforming the cell with one or more further polynucleotide sequences coding for a GST subunit, operably linked to regulatory elements that allow expression of the subunit in the cell. (As discussed above, this allows the production of heterodimeric GST dimers of the invention, or the production of homodimeric ones of the invention in greater quantities.)

Any suitable transformation method may be used, for example the transformation techniques described herein. Preferred transformation techniques include electroporation of plant protoplasts, transformation by *Agrobacterium tumefaciens* and particle bombardment. Particle bombardment is particularly preferred for transformation of monocot cells.

The cell may be in any form. for example, it may be an isolated cell, e.g. a protoplast, or it may be part of a plant tissue, e.g. a callus, or a tissue excised from a plant, or it may be part of a whole plant. Transformation may thus give rise to a chimeric tissue or plant in which some cells are transgenic and some are not.

Preferably, integration of a polynucleotide or chimeric gene of the invention into the cell's genome is achieved.

The thus obtained cell may be regenerated into a transgenic plant by techniques known in the art. These may involve the use of plant growth substances such as auxins, giberellins and/or cytokinins to stimulate the growth and/or division of the transgenic cell. Similarly,

techniques such as somatic embryogenesis and meristem culture may be used.

In many such techniques, one step is the formation of a callus, i.e. a plant tissue comprising expanding and/or dividing cells. Such calli are a further aspect of the invention as are other types of plant cell cultures and plant parts. Thus, for example, the invention provides transgenic plant tissues and parts, including embryos, meristems, seeds, shoots, roots, stems, leaves and flower parts. These may be chimeric in the sense that some of their cells are transgenic and some are not.

Regeneration procedures will typically involve the selection of transformed cells by means of marker genes. Some marker genes have already been mentioned and it should also be noted that the polynucleotides of the invention can themselves act as marker genes if they are under the control of regulatory sequences that allow their expression during the appropriate stage of the regeneration procedure. The polypeptides of the invention are capable of conferring resistance to herbicides or other phytotoxic compounds which are detoxified by GSTs on cells of the invention, as described below. Thus, an appropriate herbicide can be used to select transformants.

The regeneration step gives rise to a first generation transgenic plant. The invention also provides methods of obtaining transgenic plants of further generations this first generation plant. These are known as progeny transgenic plants. progeny plants of second, third fourth, fifth, sixth and further generations may be obtained from the first generation transgenic plant by any means known in the art.

Thus, the invention provides a method of obtaining a transgenic progeny plant comprising obtaining a second-generation transgenic progeny plant from a first-generation transgenic plant of the invention, and optionally obtaining transgenic plants of one or more further generations from the second-generation progeny plant thus obtained.

Such progeny plants are desirable because the first generation plant may not have all the characteristics required for cultivation. For example, for the production of first generation transgenic plants, a plant of a taxon that is easy to transform and regenerate may be chosen. It may therefore be necessary to introduce further characteristics in one or more subsequent generations of progeny plants before a transgenic plant more suitable for cultivation is produced.

Progeny plants may be produced form their predecessors of earlier generations by any known technique. In particular, progeny plants may be produced by:

obtaining a transgenic seed from a transgenic plant of the invention belonging to a previous generation, then obtaining a transgenic progeny plant of the invention belonging to a new generation by growing up the transgenic seed;

and/or

propagating clonally a transgenic plant of the invention belonging to a previous generation to give a transgenic progeny plant of the invention belonging to a new generation;

and/or

crossing a first-generation a transgenic plant of the invention belonging to a previous generation with another compatible plant to give a transgenic progeny plant of the invention belonging to a new generation;

and optionally;

obtaining transgenic progeny plants of one or more further generations from the progeny plant thus obtained.

These techniques may be used in any combination. for example, clonal propagation and sexual propagation may be used at different points in a process that gives rise to a transgenic plant suitable for cultivation. In particular, repetitive back-crossing with a plant taxon with agronomically desirable characteristics may be undertaken. Further steps of removing cells from a plant and regenerating new plants therefrom may also b e carried out.

Also, further desirable characteristics may be introduced by transforming the cells, plant tissues, plants or seeds, at any suitable stage in the above process, to introduce desirable coding sequences other than the polynucleotides of the invention. this may be carried out by the techniques described herein for the introduction of polynucleotides of the invention.

For example, further transgenes may be selected from those coding for other herbicide resistance traits; e.g. tolerance to Glyphosate (e.g. using an EPSP synthase gene (e.g. EP-A-0 293,358) or a glyphosate oxidoreductase (WO 92/000377) gene); or tolerance to fosametin; a dihalobenzonitrile; glufosinate (e.g. using a

phosphinotricyine acetyl transferase or glutamine synthase gene (cf. EP-A-0 242,236); asulam (e.g. using a dihydropteroate synthase gene (EP-A-0 369,367); or a sulphonylurea (e.g. using an ALS gene); diphenyl ethers such as acifluorfen or oxyfluorfen (e.g. using a protoporphyrogen oxidase gene); an oxadiazole such as oxadiazon; a cyclic imide such as chlorophthalim; a phenyl pyrrazole such as TNP, or a phenopylate or carbamate analogue thereof.

Similarly, genes for beneficial properties other than herbicide tolerance may be introduced. For example, genes for insect resistance may be introduced, notably genes encoding *Bacillus thuringiensis* (*Bt*) toxins.

Transgenic plant cells, plant parts and tissues, plants and seeds of the invention

The invention also provides transgenic plant cells, plant parts and tissues, plants and seeds. these are typically obtainable, or obtained, by the methods described above. They may be of any botanical taxon, e.g. any species or lower taxonomic grouping. Preferably, they are of a crop pant species.

Transgenic plant cells, plant parts and tissues, plants and seeds of the invention may thus be of a monocotyledonous (monocot) or dicotyledonous (dicot) taxon. Preferred dicot crop plants include tomato; potato; sugarbeet; cruciferous crops, including oilseed rape; linseed; tobacco; sunflower; fibre crops such as cotton; and leguminous crops such as peas, beans, especially soybean, and alfalfa. Preferred monocots include graminaceous plants such as wheat, maize, rice, oats, barley and rye, sorghum, triticale and sugar cane.

Wheat is particularly preferred.

Typically, a polypeptide of the invention is expressed in a plant of the invention. depending on the promoter used, this expression may be constitutive or inducible, e.g. by a herbicide safener. similarly, it may be tissue- or stage-specific, i.e. directed towards a particular plant tissue or stage in plant development.

Preferably, plant cells, plant parts and tissues, plants and seeds of the invention exhibit herbicide resistance due, at least in part, to expression of a polypeptide of the invention.

Herbicides to which plants of the invention may be resistant include Fluorodifen, Fenoxaprop-ethyl, Metolachlor, Alpha-Metolachlor, Acetochlor, Alachlor, Pretilachlor, Fluthiamid, Dimethenamid, S-Dimethenamid, Flupyrsulfuron-methyl, Triflusulfuron-methyl, Acifluorfen, Chlorimuron-ethyl, Fomesafen, Atrazine, Simazine, Cyanazine, and Metribuzin. Particularly preferred herbicides include Fenoxaprop-ethyl, Flupyrsulfuron-methyl, Fluthiamid, Acetochlor, Metolachlor and Alpha-Metolachlor. Plants of the invention may also exhibit resistance to other herbicides capable of conjugation to GSH by GSTs or to other non-herbicide phytotoxic substances.

Preferably, a transgenic plant of the invention exhibits resistance to one or more of Fenoxaprop-ethyl, Flupyrsulfuron-methyl, Fluthiamid, Acetochlor, Metalochlor and Alpha-Metolachlor. Resistance may be exhibited to herbicides which are selective for particular plant taxa and/or herbicides which are generic to all plants.

Uses of the polynucleotides, polypeptides, antibodies, probes and plants of the invention

Apart from enabling the generation of herbicide-resistant plants, the invention has a number of other uses.

Selectable markers

Polynucleotides of the invention can be used as selectable markers for detecting the transformation of plant cells. When expressed from polynucleotides of the invention, the polypeptides of the invention are capable of conferring herbicide resistance on cells of the invention, as described herein. Thus, an appropriate herbicide can be used to select transformants.

Accordingly, the invention provides a nucleic acid construct comprising:

(a) a polynucleotide of the invention operably linked to regulatory elements that allow expression of a polynucleotide of the invention a plant cell; and

(b) a site into which a further polynucleotide comprising a coding sequence can be inserted.

Preferably, site (b) is bounded by regulatory elements that allow expression of a coding sequence inserted at the site in a plant cell.

These constructs may be contained within vectors as described herein.

In these constructs, site (b) is a site into which

another nucleic acid sequence can be inserted. in cells transformed with the constructs or vectors containing them, expression of the polypeptide of the invention can be used as a selectable marker, indicating that the polynucleotide at site (b) has also been successfully introduced.

In this connection, the invention also provides a method of transforming a plant cell or of obtaining a plant cell culture or transgenic plant comprising:

(a)providing an untransformed plant cell which is susceptible to a herbicide whose herbicidal activity is reduced by a dimeric protein of the invention;

(b)transforming the plant cell with a vector comprising a marker construct of the invention;

(c)cultivating the transformed cell under conditions that allow the expression of a polypeptide of the invention;

and /or

(c')regenerating the cell to give a cell culture or plant such that a polypeptide of the invention is expressed;

and

(d)contacting the cell, cell culture or plant with the herbicide whose herbicidal activity is reduced by a dimeric protein of the invention, and to which herbicide the untransformed plant cell was susceptible;

and

(e) selecting cells, cell cultures or plants that are less susceptible to the herbicide than are corresponding untransformed cells, cell cultures or plants.

Identification of novel herbicides

The polypeptides and dimeric proteins of the invention may be used to identify compounds capable of conjugation to GSH. Thus, as conjugation to GSH is the mechanism by which GSTs are believed to effect detoxification of herbicides, the polypeptides of the invention can be used to determine whether or not a candidate herbicidal compound will be detoxified by GSTs, for example the dimeric proteins of the invention. In this case, it may be possible to develop the candidate compound as a herbicide. In particular, it may be possible to develop the candidate compound for selective use as a herbicide on crops of wheat, or of a wheat-related species, or of other plants (cf Feldman *et al supra*), having GSTs with similar activity to the dimeric proteins of the invention. This is because species having such GSTs can be expected to detoxify herbicides identified in the assay.

Accordingly, the invention provides a method of identifying compounds capable of conjugation to glutathione comprising:

(a) contacting a candidate compound suspected of being capable of being metabolised by glutathione transferase with glutathione (GSH) in the presence of a dimeric protein of the invention; and

(b) determining whether or not metabolism of the

candidate compound takes place, or to what extent takes place.

Preferably, metabolism of the compound is detected by determining whether, or to what extent, conjugation of the candidate compound to GSH takes place.

Such assay methods may be carried out by any suitable means known in the art. Compounds may be assayed singly, or, preferably, in batches containing several compounds. For example, microtitre plate-based assay techniques may be used. More specifically, the techniques of Example 3 below may be used.

The invention also provides compounds identified by the methods of the invention.

The invention also provides a kit for detecting compounds capable of being metabolised by a GST comprising:

(a) reduced glutathione or homoglutathione; and

(b) a dimeric protein of the invention.

Such kits may also comprise other components, especially buffer solutions, e.g. aqueous solutions buffered at a suitable pH (e.g. pH7 to pH10, preferably pH7 to pH8).

These kits can be used in the identification of novel herbicides.

Identification of compounds that induce GST expression

We have found that expression of the GSTs of the invention is inducible by herbicide safeners. As GSTs are

implicated in herbicide resistance, it may be desirable to identify other compounds capable of inducing their expression or that of related GSTs in wheat or other plants, preferably graminaceous plants. Such compounds may, for example, be used to induce expression of GSTs involved in herbicide tolerance. This will be beneficial as it will allow crop plants to be selectively protected from herbicides whilst weeds are killed by them.

Accordingly, the invention provides a method of identifying compounds that induce GST expression in graminaceous plants comprising:

(a) contacting a plant, preferably a graminaceous plant, or a cell or cell culture thereof, with a candidate compound suspected of being capable of inducing GST expression; and

(b) determining the level of GST expression in the plant, cell or cell culture.

Typically, the level of expression is also determined before the compound is added, or in an untreated sample, in order to provide a control. If the level of GST expression in the test sample is higher than that in the control sample then the candidate compound is an inducer of GST expression.

Preferably, the level of GST expression is determined quantitatively although, in certain situations, quantitative detection may suffice, e.g. where the level of expression is zero or undetectable in the absence of an inducer.

Determination of the level of GST expression may be performed by any suitable means. Preferably, it is

performed using antibodies or probes of the invention, as described herein.

The invention also provides compounds identified by these methods.

Antibodies that specifically recognise the polypeptides or dimeric proteins of the invention can be used to detect and preferably quantify GST expression by detecting them directly. The antibodies of the invention may thus be used for detecting polypeptides or dimeric proteins of the invention present in plant samples, e.g. by a method which comprises:

(a) providing an antibody of the invention;
(b) incubating a plant sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and
(c) determining, by any suitable technique known in the art, whether antibody-antigen complex comprising said antibody is formed.

Antibodies of the invention may be bound to a solid support and/or packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

Similarly, polynucleotides or primers of the invention or fragments thereof, labelled or unlabelled, may be used by a person skilled in the art in nucleic acid-based tests for detecting nucleic acid sequences of the invention in a sample taken from a plant, typically a wheat plant.

Such tests generally comprise bringing a sample containing DNA or RNA into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilising the probe on a solid support, removing nucleic acid in the sample which is not hybridised to the probe, and then detecting nucleic acid which has hybridised to the probe.

Alternatively, the sample nucleic acid may be immobilised on a solid support, and the amount of probe bound to such a support can be detected.

The probes of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the probe may be bound to a solid support where the assay format for which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be probed, hybridising the probe to nucleic acid in the sample, control reagents, instructions, and the like.

Measuring the level of GST in batches of seed or flour

Owing to the secondary activity of the GSTs of the invention as glutathione peroxidases, the polypeptides and dimeric proteins of the invention will also have applications in determining the quality of batches of seed and flour, especially of wheat seed, grain and wheat flour. In such batches, glutathione peroxidases are involved in reducing lipid hydroperoxides, which reduces the amount of GSH available. In particular, this occurs

during bread making. Thus, it is desirable to be able to monitor the level of GSTs having glutathione peroxidase activity in batches of seed and flour.

This can be done by any suitable means. For example, antibodies of the invention can be used to detect polypeptides or dimeric proteins of the invention, as described above. Similarly, probes of the invention can be used to detect GST mRNA, as described above.

Alternatively, to determine directly the level of GSH in a batch, the invention provides a method of determining the GSH level in a batch of seed or flour comprising:

(a) contacting a sample from the batch with a polypeptide or dimeric protein of the invention and a compound whose conjugation to GSH is catalysed by the polypeptide or protein; and

(b) determining the GSH level from the extent of reaction between the compound and GSH.

Controlling the growth of weeds

The invention also provides a method of controlling the growth of weeds at a locus where a transgenic plant of the invention is being cultivated, which method comprises applying a herbicide to the locus. Any amount of herbicide may be used, as long as it is herbicidally effective against the weeds but leaves the herbicide resistant plants of the invention unaffected, or substantially unaffected. The effect on the weeds may be, for example, to kill them or to inhibit their growth.

Any type of weed that responds to a particular herbicide may be controlled in this way. *Alopecurus*

myosuroides, Avena fatua, Lolium spp., Bromus spp., Poa annua, Galium aparine, Aper spica-venti, Matricaria inodora, Stellaria media, Papaver rholas, Polygonum spp., Setaria sp., Sorghum halapense, Panicum miliaceum, Echinochloa spp., Digitaria sanguinalis, Phalaris minor, Abutilon theophrasti, Amaranthus retroflexus, chenopodium album, Datura stramoniuon, Solanum nigrum, Xanthium strumarium, saggitaria spp., Monochoria vaginalis, Lindernia spp., Eleokaris kurogaai, Scirpus juncoides, Cyperus spp.

The herbicide may, for example, be one whose activity is identified by the methods of the invention (see above). Alternatively, it may be a known herbicide, for example one of the herbicides mentioned herein.

The herbicide may be applied at any suitable time during the life cycle of the transgenic plant, for example pre-emergence or post-emergence. Timing of application will be tailored to the development of the weeds which it is desired to control. Where inducible or tissue- and/or stage- specific expression of the active dimer of the invention is employed, timing of herbicide application will be tailored to the optimal expression of the invention in the course of the development of the transgenic plant of the invention.

The following Examples illustrate the invention.

EXAMPLES

Example 1: Isolation and characterisation of the nucleotide sequence encoding *Ta*GST1

(a) Purification of wheat GST isoenzymes

Wheat GST isoenzymes containing the *Ta*GST1 subunit were purified by the method of Dixon *et al* (Pestic. Sci. 1997, 50, 72-82). This is summarised below.

Wheat seeds (*Triticum aestivum* L. var. Hunter) were imbibed in a 10 mg/l solution of the herbicide safener fenchlorazole-ethyl and then grown in an environmental growth room with further root-applied watering treatments of 5 mg/l fenchlorazole ethyl applied as required. At 10 days after imbibing, the shoot tissue was harvested and extracted prior to precipitation of the protein with ammonium sulphate (80% saturation). The total protein extract was then applied in the presence of 1 M ammonium sulphate to a phenyl-Sepharose column. The bound GSTs were then recovered, firstly by reducing the ammonium sulphate concentration to 0 M to yield the polar GST fraction, which represented 61% of the recovered activity toward 1-chloro-2,4-dinitrobenzene (CDNB). The remaining 39% of the GST activity was then recovered by adding ethylene glycol (50 % v/v) to the running buffer to yield the hydrophobic GST fraction.

The polar and hydrophobic GST fractions were then independently applied to the affinity matrix, *S*-hexyl-glutathione agarose. This matrix bound 90% of the GST activity toward CDNB. Prior to elution of the column with the ligand, *S*-hexyl-glutathione, the matrix was washed with phosphate buffer, followed by phosphate buffer containing 200 mM potassium chloride. The GSTs eluting in this salt wash were termed the "loosely-bound" fraction. Tightly-bound proteins were then eluted with 5 mM *S*-hexyl-glutathione. With both the polar and hydrophobic GSTs an average of 34% of the GST activity toward CDNB eluted in the loosely-bound fraction and 66% eluted in the presence of *S*-hexyl-glutathione. The loosely-bound fraction contained the GSTs which will be considered in

Example 2. The major wheat GSTs of interest in this example were found in the affinity-purified pool and to define the numbers of isoenzymes and component subunits present, this pool was analysed in detail.

When the affinity-bound pools of the polar and hydrophobic GSTs were analysed by anion-exchange chromatography on Q-sepharose, the partial resolution of the eluting activity suggested the presence of multiple isoenzymes (Figure 1). The component polypeptides in the active fractions were then analysed by silver staining after resolution by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). It was concluded that in fenchlorazole-ethyl-treated wheat the polar GSTs were composed of 25 kDa and 26 kDa polypeptides, while the hydrophobic fraction contained 25kDa, 26 kDa and 24 kDa polypeptides. Further analysis by reversed-phase HPLC confirmed the subunit compositions (Figure 2). Based on the combined analyses by Q-sepharose, HPLC and SDS-PAGE these GST polypeptides were named as described in Table 1, which also contains details of how these subunits associate together to form the active dimers found in plants and the relative abundance of these subunits in extracts from unsafened and fenchlorazole-ethyl treated (safened) plants.

(b) GST activities of the purified *Ta*GST isoenzymes

The purified isoenzymes were assayed for GST activity toward herbicides using the HPLC-based assays described by Edwards R. and Cole D.J. (Pesticide Biochemistry and Physiology Vol. 54, pp96-104 (1996)) and the results are presented in Table 2. Both polar and hydrophobic GSTs from the affinity-bound pools of isoenzymes showed detoxifying activities toward the

selective graminicide fenoxaprop-ethyl, the diphenyl-ether herbicide fluorodifen, and the chloracetanilide metolachlor. These isoenzymes had additional activities as glutathione peroxidases able to reduce linoleic acid hydroperoxide, a major reaction product formed during membrane peroxidation in plants (Williamson and Beverly. J. Cereal Sci. 8, 1988, 155-163).

(c) Preparation of polyclonal antibodies to the major wheat GST isoenzymes

Purified *Ta*GST1-1 was used to immunise rabbits to raise polyclonal antibodies to the differing isoenzymes. The reactivity of the antiserum toward crude wheat preparations was demonstrated with a Western blot of polypeptides resolved by SDS-PAGE. The antibodies were then used to identify the corresponding cDNAs in an expression library.

(d) Identification and characterisation of a cDNA encoding *Ta*GST1

An expression library was prepared from poly (A)+ RNA extracted from 7-day wheat shoots grown from seed treated with fenchlorazole-ethyl. The library was constructed in lambda ZAP II (Stratagene) and plaque forming units (pfus) screened with the antiserum raised against *Ta*GST1-1. From an initial screen of 170,000 pfus 17 positive plaques were identified, of which 12 were further purified to homogeneity in secondary and tertiary screens and the wheat cDNAs excised from the phage to form Bluescript plasmids in *E. coli* SOLR. (Stratagene). Automated DNA sequencing showed that all clones had an identical coding sequence, although differences in the 5' and 3' untranslated regions were apparent, such that of 6

clones sequenced fully on both strands, 4 different untranslated regions were observed. Since these clones shared a common open reading frame they were all designated *TaGST1* and then subdivided as A, B, C and D. The nucleotide sequence of *TaGST1* showing the variable untranslated regions of A, B, C and D is shown in SEQ ID No. 1, together with the deduced amino acid sequence of the coding region (SEQ ID No. 2).

To confirm that *TaGST1* encoded a GST, it was expressed as a fusion protein with beta-galactosidase using the pBluescript plasmid in *E. coli* strain SOLR. *TaGST1* clones were inoculated into LB liquid medium and were grown overnight at 37°C on an orbital shaker in the presence of IPTG. Bacteria were then pelleted by centrifugation, lysed by sonication and assayed for GST activity toward CDNB and analysed by SDS-PAGE and Western blotting using the anti-*Ta*GST1-1 serum. With all six *TaGST1* clones, GST activity could be determined in the crude extracts in the range 30 - 50 nkat/ mg crude lysate. This was in contrast to control *E. coli* containing the bluescript plasmid without a cDNA insert which showed negligible GST activity (0.2 nkat/mg). When the polypeptides contained in the lysates of the various *Ta*GST1 clones were analysed by SDS-PAGE, in every case the *Ta*GST1-fusion protein was clearly visible as a highly expressed polypeptide relative to the controls. The differences in size of the fusion proteins resulted from variations in the length of 5' untranslated sequence from the *Ta*GST1 included in the fusion. All the fusion proteins reacted with the anti-*Ta*GST1 serum.

To confirm that the GST activity in the extracts from *Ta*GST1 clones was due to the fusion protein, the

GST-fusion was purified from the clone *Ta*GST1.2 using *S*-hexyl-glutathione agarose affinity chromatography. The pure fusion protein was then assayed for enzyme activity toward herbicide and hydroperoxide substrates and was found to show a similar spectrum of activities to that of the pure *Ta*GST1-1 isoenzyme from wheat shoots.

Table 1

Summary of the characteristics of major classes of wheat GST isoenzymes.

The GST subunits had the following retention times by reversed-phase HPLC. *Ta*GST1a - 26.4 min, *Ta*GST1b - 27.1 min, *Ta*GST2 - 31.1 min, *Ta*GST3 - 30.9 min, *Ta*GST4 - 33.2 min.

Isoenzyme type	Subunits	Polar (P) or Hydrophobic (H)	Molecular weight (kDa)	Anti-*Ta*GST1 antibody reaction	% Enhancement by safener
*Ta*GST1-1	*Ta*GST1a	P	25	+	30-50
	*Ta*GST1b	P	25	+	30-50
*Ta*GST1-2	*Ta*GST1a	P	25	+	Only observed with safener
	*Ta*GST1b	P	25	+	
	*Ta*GST2	P	26	-	
*Ta*GST1-3	*Ta*GST1a	P	25	+	Only observed with safener
	*Ta*GST1b	P	25	+	
	*Ta*GST3	H	26	-	
*Ta*GST1-4	*Ta*GST1a	P	25	+	300%
	*Ta*GST1b	P	25	+	300%
	*Ta*GST4	H	24	-	300%

Table 2

Activity of GST isoenzymes purified from fenchlorazole-ethyl-treated wheat shoots.

Enzyme activities are expressed as nkat.mg-1

Isoenzyme	CDNB	Fluorodifen	Fenoxaprop-ethyl	Metolachlor
Polar				
*Ta*GST1-1	1,528	0.97	0	0.11
*Ta*GST1-2	1,441	0.38	0.61	0.25
Hydrophobic				
*Ta*GST1-3	1,700	0.38	0.44	0.28
*Ta*GST1-4	1,553	0.57	0.23	0.23

Table 3

CDNB and herbicide activities of recombinant wheat GSTs

Activities expressed as nkat.mg-1 ± standard error

Recombinant enzyme	CDNB	Fluorodifen	Fenoxaprop-ethyl	Metolachlor
*Ta*GST1	1970 ± 30	2.0 ± 0.1	0	0.127 ± 0.014
WIC 1	406.5 ± 0.5	0.136 ± 0.011	0.050 ± 0.010	0.315 ± 0.003
WIC 2	187 ± 1	0.096 ± 0.002	0.085 ± 0.002	0.512 ± 0.04
WIC 3	2,519 ± 88	0.014 ± 0.006	0.093 ± 0.002	0.053 ± 0.004
WIC 4	980 ± 86	0.036 ± 0.004	0.012 ± 0.001	0.037 ± 0.003
WIC 5	174 ± 8	0.030 ± 0.002	0.067 ± 0.003	0.040 ± 0.004
TA 27	237 ± 13	0.034 ± 0.003	0.036 ± 0.004	0.063 ± 0.006

EXAMPLE 2: Cloning of wheat GSTs resembling the type 1 GSTs from maize.

(a) Characterisation of type 1 GSTs in wheat

The observation that extracts from safener-treated wheat shoots contained GSTs which, unlike those described in Example 1, were not selectively retained on the affinity matrix suggested that a discrete class of GSTs were present in this loosely bound fraction. Crude extracts of wheat seedlings were analysed by Western blotting following SDS-PAGE using a polyclonal rabbit antiserum raised to the *Zm*GSTI-II heterodimer. The antiserum reacted strongly with several polypeptides of Mr 23 - 27 kDa. These polypeptides were present in the loosely-bound fraction from the *S*-hexyl-glutathione affinity column, but not in the affinity bound fraction.

(b) Cloning of cDNAs from a wheat expression library

Having established that safener-treated wheat shoots contained polypeptides which cross-reacted with the antiserum raised to the maize GSTs, the primary cDNA expression library prepared from fenchlorazole-ethyl treated wheat shoots was screened with the antibody for positive clones. Following a screen of 170,000 pfu., ten positive plaques were identified, with obvious differences in the intensity of recognition, with four plaques showing a strong colour reaction and six plaques of lower intensity. These cDNA clones were termed *WIC* clones. All four of the stronger-reacting plaques (*WIC* 1, 2, 4 and 5) and four of the weaker positives (WIC 3, 7, 8 and 10) were purified to homogeneity, the respective plasmids excised and DNA preparations sequenced. The clones were then grouped by their degree of similarity in sequence.

In the sequence listing, WIC 1 is SEQ ID No. 3 and its deduced amino acid sequence is SEQ ID No. 4. WIC 2 is SEQ ID No. 5 and its deduced amino acid sequence is SEQ ID No. 6. The coding sequences of WIC 3, WIC 7 and WIC 8 were identical in sequence. The DNA sequence of WIC 3/7/8 is given in SEQ ID No. 7 and the deduced amino acid sequence in SEQ ID No. 8 All three sequences contained a stop codon in the 5' untranslated region of the GST gene, although some expression occurred. The DNA sequence of WIC 5 is shown in SEQ ID No. 9, and the deduced amino acid sequence in SEQ ID No. 10. WIC 4 and WIC 10 had identical coding sequences, but differed in their untranslated regions. In particular, WIC 10 had a stop codon in the 5' untranslated region, though this did not prevent all expression of the fusion protein The WIC 4 DNA sequence is given in SEQ ID No. 11 and the deduced WIC 4/10 amino acid sequence in SEQ ID No. 12 (the WIC 10 DNA sequence is not shown).

(c) Cloning of wheat GSTs by differential screening of a cDNA library

A further cDNA clone, termed *TA 27* was obtained. A cDNA library prepared from wheat seedlings treated with the herbicide safener cloquintocet-mexyl, was screened for clones which represented mRNAs which were differentially expressed in wheat in response to safener application. The identity of the clone as a GST was suggested from its nucleotide (SEQ ID No. 13) and deduced amino acid (SEQ ID No. 14) sequence. As the coding sequence of *TA* 27 was not in frame with beta-galactosidase in the pBluescript vector, the coding sequence was sub-cloned into the expression vector pET 11a (Novagen), with translation starting at the first ATG codon in the clone, which gave a reasonable alignment of the open reading frame with that of other GSTs involved

in herbicide metabolism, notably the *Zm*GSTIV sequence.

(d) Activity of recombinant GSTs of the invention

To confirm that the *WIC* clones and *TA 27* encoded functional GSTs the corresponding enzymes were expressed as recombinant enzymes in *E. coli*. The full coding sequence of *TA 27* was expressed in the pET vector, while the WIC clones were expressed as fusions with part of the beta-galactosidase enzyme using the pBluescript vector. The levels of recombinant protein expressed varied between the differing clones. Appreciable amounts of recombinant protein were observed in the *TA 27* pET clones and in clones *WIC 1, WIC 2, WIC 4* and *WIC 5*. Western blotting of these total bacterial extracts with the antiserum raised to *Zm*GSTI-II showed that the fusion proteins were selectively recognised by the antiserum. In contrast, use of the antiserum demonstrated much lower levels of expression of immunoreactive fusion proteins in clones *WIC 3, WIC 7, WIC 8* and *WIC 10*.

To assay the recombinant fusion proteins for GST activity , the *E. coli* were grown in the presence of IPTG and then pelleted by centrifugation. The bacteria were then lysed by sonication and the protein precipitated using 80% ammonium sulphate. After resuspension and desalting, GSTs were purified by affinity chromatography. The WIC 3 fusion protein was purified using sulphobromophthalein-*S*-glutathione affinity chromatography (Mozer *et al*. Biochem. 22, 1983, 1068-1072) while the other WIC fusion proteins were purified using glutathione-agarose (Mannervik and Guthenberg. Methods Enzymol. 77, 1981, 231-235).The purified enzymes were then assayed for GST activities toward herbicides (Table 3) and GST activities toward non-herbicide

substrates and glutathione peroxidase activities toward organic hydroperoxides (Table 4).

Table 4

Other GST activities and glutathione peroxidase activities.

Activities expressed as nkat.mg-1 ± standard error. Peroxidase activities expressed as absorbance change at 340 nm.mg-1 ± standard error (n=3). N.D = not detected.

	Cumene Hydroperoxide	Benzyl Isothiocyanate	Crotonaldehyde	ethacrynic acid
WIC 1	18.6 ± 0.5	18.0 ± 3.75	7.1 ± 0.7	N.D.
WIC 2	28.2 ± 1.7	33.3 ± 4.5	5.5 ± 1.3	N.D.
WIC 3	1.4 ± 0.3	9.0 ± 0.5	6.3 ± 0.9	N.D.
WIC 4	6.2 ± 0.3	4.2 ± 0.4	5.5 ± 0.6	1.4 ± 0.3
WIC 5	1.3 ± 0.2	9.4 ± 2.0	4.5 ± 0.5	N.D.
IC GST 25	0.7 ± 0.1	11.8 ± 0	3.7 ± 0.3	N.D.
TA 27	peroxidase activity as WIC 1 and WIC 2			

EXAMPLE 3: A microtitre plate - based screen to identify herbicidal molecules which are metabolised by GSTs of the invention and may selectively control weeds in a crop of wheat or other species such as maize, soybean or rice

(a) Degradation of candidate herbicides by wheat GSTs and relationship to crop and weed selectivity

Herbicidal molecules which are degraded by recombinant wheat GSTs may be predicted to be tolerated by plants of wheat or other crop species. These herbicides may be less rapidly degraded in weeds such as black-grass (*Alopecurus myosuroides*) that it is desirable to control in a crop of wheat or other species. Herbicides found in a laboratory based screen to be metabolised by these GSTs are therefore likely to possess useful abilities to selectively control troublesome weeds in a crop of wheat, or other species such as maize, soybean, rice, cotton, barley, oat, rye, sorhum, triticale, potato, sugarcane or sugarbeet.

(b) A 96 well plate - based assay procedure for identifying novel herbicides degraded by recombinant wheat GSTs

Novel herbicides arising from a chemical synthesis programme oriented to optimisation for selective herbicidal activity and potency may be screened for ability to be degraded by a panel of recombinant GSTs using a 96 well microplate assay format and subsequent reaction analysis by automated High Pressure Liquid Chromatography (HPLC). This allows for example, the screening of a set of eleven novel herbicides and one positive control compound such as CDNB, against a panel of seven recombinant GSTs. An eighth file of wells

contains test compounds but lacks GSTs; these wells serve to identify non-enzymic reaction of the test compounds with reduced glutathione. Alternatively, the array can be configured to screen more test compounds against a more limited number of GSTs. For example, fifteen compounds can be screened against five GSTs or forty seven compounds may be screened with a single mixture of GSTs. In all cases, provision is made for a positive control and to test for non-enzymic reaction with reduced glutathione.

Enzyme assays are carried out in a total reaction volume of 100 microlitres. Each reaction mixture contains 100 micromolar Tris.HCl buffer, pH 7.8, 500 micromolar reduced glutathione and where appropriate, 500 micromolar test compound or a reference substrate such as CDNB; and 14 micrograms of GST protein. The microplate is incubated at 30°C on a variable speed agitator for 30 minutes and reactions are then stopped by the addition of 15 microlitres of 23% perchloric acid solution. The microplate is then centrifuged at 2000 g for 15 minutes.

(c) Reaction analysis by automated High Pressure Liquid Chromatography.

The separation and analysis of glutathione conjugates of test herbicides may be carried out using High Pressure Liquid Chromatography (HPLC), for example a Gilson HPLC in tandem with corresponding software, for example Gilson Version 7.12 and fitted with an appropriate column, for example a 5 cm Spherisorb ODS2 column. Typically, separation may be carried out using a two phase solvent system as follows: Phase A: water containing 0.1% trifluoroacetic acid and 5.5% acetonitrile; Phase B: 100% acetonitrile; flow rate 1.5 ml per minute; injection volume 20 microlitres.

The elution gradient may be typically as follows: 10% phase B for one minute, followed by a linear gradient to reach 60% phase B after 8.5 minutes. The gradient is further increased to reach 100% phase B at 9 minutes; phase B is continued at 100% until 11.5 minutes and is then reduced in a linear gradient to 10% at 13.5 minutes. A further 1.5 minutes at 10% phase B is required to re-equilibrate the column. Absorbance signals are detected at 264 nanometres using a suitable UV detector, and detect the glutathione conjugate of CDNB, having a retention time of 2.4 minutes, resolving this from unreacted CDNB having a retention time of 4.75 minutes. Such conditions also allow for the resolution and detection of the glutathione conjugates arising from the metabolism of other reference herbicides such as metolachlor, fenoxaprop, fenoxaprop-ethyl and fluorodifen and also of a variety of novel herbicidal compounds identified in the assay.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: RHONE-POULENC AGRICULTURE LIMITED
(B) STREET: FYFIELD ROAD
(C) CITY: ONGAR
(D) STATE: ESSEX
(E) COUNTRY: UNITED KINGDOM
(F) POSTAL CODE (ZIP): CM5 0HW

(ii) TITLE OF INVENTION: GLUTATHIONE TRANSFERASES

(iii) NUMBER OF SEQUENCES: 14

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1085 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:46..711

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..1085
(D) OTHER INFORMATION:/note= "SEQUENCE OF *Ta*GST1 AND ENCODED AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CAAACACAAG CACAGATCGG TCGAGATTCA AGGCAACCGG GAGCA ATG GCG GGC          54
                                                  Met Ala Gly

GAG AAG GGG CTG GTG CTG CTG GAC TTC TGG GTG AGC CCG TTC GGG CAG       102
Glu Lys Gly Leu Val Leu Leu Asp Phe Trp Val Ser Pro Phe Gly Gln
```

```
CGC GTG CGC ATC GCG CTG GCC GAG AAG GGC CTG CCC TAC GAG TAC GCG      150
Arg Val Arg Ile Ala Leu Ala Glu Lys Gly Leu Pro Tyr Glu Tyr Ala

GAG GAG GAC CTG ATG GCC GGC AAG AGC GAC CGC CTC CTC CGC GCC AAC      198
Glu Glu Asp Leu Met Ala Gly Lys Ser Asp Arg Leu Leu Arg Ala Asn

CCG GTG CAT AAG AAG ATC CCG GTG CTC CTC CAC GAC GGC CGT GCC GTC      246
Pro Val His Lys Lys Ile Pro Val Leu Leu His Asp Gly Arg Ala Val

AAC GAG TCC CTC ATC ATC CTC CAG TAC CTG GAG GAG GCC TTC CCG GAC      294
Asn Glu Ser Leu Ile Ile Leu Gln Tyr Leu Glu Glu Ala Phe Pro Asp

GCG CCC GCT CTG CTC CCC TCC GAC CCC TAC GCG CGC GCG CAG GCC CGC      342
Ala Pro Ala Leu Leu Pro Ser Asp Pro Tyr Ala Arg Ala Gln Ala Arg

TTC TGG GCC GAC TAC GTC GAC AAG AAG GTC TAC GAC TGC GGC TCC CGC      390
Phe Trp Ala Asp Tyr Val Asp Lys Lys Val Tyr Asp Cys Gly Ser Arg

CTC TGG AAG CTC AAG GGC GAG CCG CAG GCG CAG GCG CGC GCC GAG ATG      438
Leu Trp Lys Leu Lys Gly Glu Pro Gln Ala Gln Ala Arg Ala Glu Met

CTG GAC ATC CTC AAG ACC CTC GAC GGC GCG CTC GGG GAC AAG CCC TTC      486
Leu Asp Ile Leu Lys Thr Leu Asp Gly Ala Leu Gly Asp Lys Pro Phe

TTC GGC GGC GAC AAG TTC GGG TTC GTC GAC GCC GCC TTC GCG CCC TTC      534
Phe Gly Gly Asp Lys Phe Gly Phe Val Asp Ala Ala Phe Ala Pro Phe

ACC GCG TGG TTC CAC AGC TAC GAG AGG TAC GGC GAG TTC AGC CTG CCG      582
Thr Ala Trp Phe His Ser Tyr Glu Arg Tyr Gly Glu Phe Ser Leu Pro

GAG GTG GCG CCC AAG ATC GCC GCG TGG GCC AAG CGC TGC GGC GAG CGG      630
Glu Val Ala Pro Lys Ile Ala Ala Trp Ala Lys Arg Cys Gly Glu Arg

GAG AGC GTC GCC AAG AGC CTC TAC TCG CCG GAC AAG GTG TAC GAC TTC      678
Glu Ser Val Ala Lys Ser Leu Tyr Ser Pro Asp Lys Val Tyr Asp Phe

ATC GGC CTG CTC AAG AAG AAG TAC GGC ATC GAG TA GGCGCGCCGA            723
Ile Gly Leu Leu Lys Lys Lys Tyr Gly Ile Glu

CGGACGGACG GACGGGCCAT GCAGGCGACA GCCGGCCCGC CGTCCGGAGG GAAGCAACAA    783

ATAAATCAGG GAGCGATTTG GGTGGCCTAC AATGCGTACG TCTGGATAGA GTATTTCTTT    843

CTTTCTTTCT TCGTGGAATA AAGTGCTCCG TGTGTGTGTG GTTGGTGGTT GTTGGTTGGA    903

TCAGTCAGTG TGTGTGGGTG CGTGTTGTGT ACTCAGTACT CGTGATGTGT GTGTGTGTCA    963

ATGTGTCAAC CCTGGTCTTC GGTGGGGGCA GCACCGAGTT GCCACCTGCC ATTCCATTTC   1023

CATTCCGGGC GATGAATAAA TTAAAAAAGA GTCTCATTTG TTTAAAAAAA AAAAAAAAAA   1083

AA                                                                  1085
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ala Gly Glu Lys Gly Leu Val Leu Leu Asp Phe Trp Val Ser Pro
 1               5                  10                  15

Phe Gly Gln Arg Val Arg Ile Ala Leu Ala Glu Lys Gly Leu Pro Tyr
            20                  25                  30

Glu Tyr Ala Glu Glu Asp Leu Met Ala Gly Lys Ser Asp Arg Leu Leu
        35                  40                  45

Arg Ala Asn Pro Val His Lys Lys Ile Pro Val Leu Leu His Asp Gly
    50                  55                  60

Arg Ala Val Asn Glu Ser Leu Ile Ile Leu Gln Tyr Leu Glu Glu Ala
65                  70                  75                  80

Phe Pro Asp Ala Pro Ala Leu Leu Pro Ser Asp Pro Tyr Ala Arg Ala
                85                  90                  95

Gln Ala Arg Phe Trp Ala Asp Tyr Val Asp Lys Lys Val Tyr Asp Cys
            100                 105                 110

Gly Ser Arg Leu Trp Lys Leu Lys Gly Glu Pro Gln Ala Gln Ala Arg
        115                 120                 125

Ala Glu Met Leu Asp Ile Leu Lys Thr Leu Asp Gly Ala Leu Gly Asp
    130                 135                 140

Lys Pro Phe Phe Gly Gly Asp Lys Phe Gly Phe Val Asp Ala Ala Phe
145                 150                 155                 160

Ala Pro Phe Thr Ala Trp Phe His Ser Tyr Glu Arg Tyr Gly Glu Phe
                165                 170                 175

Ser Leu Pro Glu Val Ala Pro Lys Ile Ala Ala Trp Ala Lys Arg Cys
            180                 185                 190

Gly Glu Arg Glu Ser Val Ala Lys Ser Leu Tyr Ser Pro Asp Lys Val
        195                 200                 205

Tyr Asp Phe Ile Gly Leu Leu Lys Lys Lys Tyr Gly Ile Glu
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 865 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:54..725

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..865
(D) OTHER INFORMATION:/note= "WIC1 SEQUENCE AND ENCODED IC1 AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
GGAACTCAAC CATTGATCTT CAAGAAGCGG AAGCAAACAG AGCAAAAGGT GTG ATG          56
                                                              Met

GCG GCG CCG GCG GTG AAG GTG TAC GGG TGG GCG ATG TCG CCG TTC GTG          104
Ala Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Met Ser Pro Phe Val

GCG CGC GCG CTG CTG TGC CTG GAG GAG GCC GGC GTG GAG TAC GAG CTC          152
Ala Arg Ala Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu Leu

GTC CCC ATG AGC CGC GAG GCC GGC GAC CAC CGC CAG CCC GAC TTC CTC          200
Val Pro Met Ser Arg Glu Ala Gly Asp His Arg Gln Pro Asp Phe Leu

GCC CGG AAC CCC TTC GGC CAG GTC CCC GTT CTC GAG GAC GGC GAC CTC          248
Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp Leu

ACC ATC TTC GAG TCG CGC GCC GTC GCG AGG CAC GTG CTG CGC AAG CAC          296
Thr Ile Phe Glu Ser Arg Ala Val Ala Arg His Val Leu Arg Lys His

AAA CCG GAG CTG CTG GGC TCC GGC TCG CCG GAG TCG GCG GCG ATG GTG          344
Lys Pro Glu Leu Leu Gly Ser Gly Ser Pro Glu Ser Ala Ala Met Val

GAC GTG TGG CTG GAG GTG GAG GCC CAC CAG CAC CAG ACC CCG GCG GGC          392
Asp Val Trp Leu Glu Val Glu Ala His Gln His Gln Thr Pro Ala Gly

ACC ATC GTC ATG CAG TGC ATC CTC ACC CCG TTC CTC GGC TGC CAG CGC          440
Thr Ile Val Met Gln Cys Ile Leu Thr Pro Phe Leu Gly Cys Gln Arg

GAC CAG GCC GCC ATC GAC GAG AAC GCG GCA AAG CTG ACG AAT CTG TTC          488
Asp Gln Ala Ala Ile Asp Glu Asn Ala Ala Lys Leu Thr Asn Leu Phe

GAC GTG TAC GAG GCG CGC CTG TCG GCG TCG AGG TAC CTT GCC GGG GAG          536
Asp Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly Glu

GCG GTC AGC CTC GCG GAC CTC AGC CAC TTC CCG TTC ATG CGA TAC TTC          584
```

```
Ala Val Ser Leu Ala Asp Leu Ser His Phe Pro Phe Met Arg Tyr Phe

ATG GAC ACC GAG TAC GCG TCG CTG GTG GAG GAG CGC CCG CAC GTG AAG      632
Met Asp Thr Glu Tyr Ala Ser Leu Val Glu Glu Arg Pro His Val Lys

GCG TGG TGG GAG GAG TTC AAG GCC AGC CCG GCG GCG AAG AGG GTG ACG      680
Ala Trp Trp Glu Glu Phe Lys Ala Ser Pro Ala Ala Lys Arg Val Thr

GAG TTC ATG CCG CCA AAC TTC GGG TTC GGA AAG AAG GCA GAG AAG          725
Glu Phe Met Pro Pro Asn Phe Gly Phe Gly Lys Lys Ala Glu Lys

TGATGACAAG AACGAACACC GAGCGAACAT GTTGTGTGGT CTGTGCGACC CGACCATGGC   785

TCAATGTTTT GGGCTGTTTG TGTTTCACGC ATGAATGAAT AAAACAAAAT GCTTTTGGGT   845

TTCAAAAAAA AAAAAAAAAA                                               865
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 224 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Ala Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Met Ser Pro Phe
 1               5                  10                  15

Val Ala Arg Ala Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu
            20                  25                  30

Leu Val Pro Met Ser Arg Glu Ala Gly Asp His Arg Gln Pro Asp Phe
        35                  40                  45

Leu Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp
    50                  55                  60

Leu Thr Ile Phe Glu Ser Arg Ala Val Ala Arg His Val Leu Arg Lys
65                  70                  75                  80

His Lys Pro Glu Leu Leu Gly Ser Gly Ser Pro Glu Ser Ala Ala Met
                85                  90                  95

Val Asp Val Trp Leu Glu Val Glu Ala His Gln His Gln Thr Pro Ala
            100                 105                 110

Gly Thr Ile Val Met Gln Cys Ile Leu Thr Pro Phe Leu Gly Cys Gln
        115                 120                 125

Arg Asp Gln Ala Ala Ile Asp Glu Asn Ala Ala Lys Leu Thr Asn Leu
    130                 135                 140
```

```
Phe Asp Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly
145                 150                 155                 160

Glu Ala Val Ser Leu Ala Asp Leu Ser His Phe Pro Phe Met Arg Tyr
                165                 170                 175

Phe Met Asp Thr Glu Tyr Ala Ser Leu Val Glu Glu Arg Pro His Val
            180                 185                 190

Lys Ala Trp Trp Glu Glu Phe Lys Ala Ser Pro Ala Ala Lys Arg Val
        195                 200                 205

Thr Glu Phe Met Pro Pro Asn Phe Gly Phe Gly Lys Lys Ala Glu Lys
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 930 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:60..725

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..930
(D) OTHER INFORMATION:/note= "WIC2 SEQUENCE AND ENCODED IC2 AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
CACGCGTCCA TCTCCAAGAA GCGGAAGCTA GTGGAGCAGA GCAAACCAAG CAAGGTTGG        59

ATG GCG CCG GCG GTG AAG GTG TAC GGG TGG GCC GTG TCG CCG TTC GTG        107
Met Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Val Ser Pro Phe Val

GCG CGC CCA CTG CTG TGC CTG GAG GAG GCC GGC GTC GAG TAC GAG CTC        155
Ala Arg Pro Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu Leu

GTG TCC ATG AGC CGC GCG GCC GGC GAC CAC CGC CAG CCG GAC TTC CTC        203
Val Ser Met Ser Arg Ala Ala Gly Asp His Arg Gln Pro Asp Phe Leu

GCC CGG AAC CCC TTC GGC CAG GTC CCC GTC CTC GAG GAC GGC GAC CTC        251
Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp Leu

ACC CTC TTC GAG TCG CGC GCG ATC GCG AGG CAC GTG CTC CGG AAG CAC        299
```

```
Thr Leu Phe Glu Ser Arg Ala Ile Ala Arg His Val Leu Arg Lys His

AAG CCG GAG CTG CTG GGC TGC GGC TCG CCG GAG GCG GAG GCG ATG GTG      347
Lys Pro Glu Leu Leu Gly Cys Gly Ser Pro Glu Ala Glu Ala Met Val

GAC GTG TGG CTG GAG GTG GAG GCC CAC CAG TAC AAC CCC GCG GCC AGC      395
Asp Val Trp Leu Glu Val Glu Ala His Gln Tyr Asn Pro Ala Ala Ser

GCC ATC GTG GTG CAG TGC ATC ATC TTG CCG CTA CTG GGC GGC GCG CGG      443
Ala Ile Val Val Gln Cys Ile Ile Leu Pro Leu Leu Gly Gly Ala Arg

GAC CAG GCG GTG GTG GAC GAG AAC GTA GCC AAG CTC AAG AAG GTG CTG      491
Asp Gln Ala Val Val Asp Glu Asn Val Ala Lys Leu Lys Lys Val Leu

GAG GTG TAC GAG GCA CGG CTG TCG GCG TCC AGG TAC CTC GCC GGG GAC      539
Glu Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly Asp

GAC ATC AGC CTC GCC GAC CTC AGC CAC TTC CCC TTC ACG CGC TAC TTC      587
Asp Ile Ser Leu Ala Asp Leu Ser His Phe Pro Phe Thr Arg Tyr Phe

ATG GAG ACG GAG TAC GCG CCG CTG GTG GCG GAG CTC CCC CAC GTG AAC      635
Met Glu Thr Glu Tyr Ala Pro Leu Val Ala Glu Leu Pro His Val Asn

GCG TGG TGG GAG GGG CTC AAG GCC AGG CCG GCC GCG AGG AAG GTG ACG      683
Ala Trp Trp Glu Gly Leu Lys Ala Arg Pro Ala Ala Arg Lys Val Thr

GAG CTC ATG CCG CCG GAC CTT GGG CTT GGA AAG AAA GCA GAG              725
Glu Leu Met Pro Pro Asp Leu Gly Leu Gly Lys Lys Ala Glu

TAGTGATGAC TGCCGCCAAC GTTCACCAGG ATCGAGCAAG TCACTGTCGA GTCTCCGGTT    785

TTGCGTTGTA CGGCACCGGG GCACCGGCCT ATATTTTCTG TACCAGTGGC TCGTGTTTTG    845

ATGTTTTAGT CTCACGCTTG AATAAAATGC AAGATATACC CATCGGTTCT AAAAGAAAAA    905

AAAAAAAAAA AAAAAAAAAA AAAAA                                          930
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Val Ser Pro Phe Val
 1               5                  10                  15

Ala Arg Pro Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu Leu
            20                  25                  30

Val Ser Met Ser Arg Ala Ala Gly Asp His Arg Gln Pro Asp Phe Leu
```

```
         35                  40                  45

Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp Leu
     50                  55                  60

Thr Leu Phe Glu Ser Arg Ala Ile Ala Arg His Val Leu Arg Lys His
 65                  70                  75                  80

Lys Pro Glu Leu Leu Gly Cys Gly Ser Pro Glu Ala Glu Ala Met Val
                 85                  90                  95

Asp Val Trp Leu Glu Val Glu Ala His Gln Tyr Asn Pro Ala Ala Ser
            100                 105                 110

Ala Ile Val Val Gln Cys Ile Ile Leu Pro Leu Leu Gly Gly Ala Arg
        115                 120                 125

Asp Gln Ala Val Val Asp Glu Asn Val Ala Lys Leu Lys Lys Val Leu
    130                 135                 140

Glu Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly Asp
145                 150                 155                 160

Asp Ile Ser Leu Ala Asp Leu Ser His Phe Pro Phe Thr Arg Tyr Phe
                165                 170                 175

Met Glu Thr Glu Tyr Ala Pro Leu Val Ala Glu Leu Pro His Val Asn
            180                 185                 190

Ala Trp Trp Glu Gly Leu Lys Ala Arg Pro Ala Ala Arg Lys Val Thr
        195                 200                 205

Glu Leu Met Pro Pro Asp Leu Gly Leu Gly Lys Lys Ala Glu
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 927 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:72..707

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..927
(D) OTHER INFORMATION:/note= "WIC 3/7/8 SEQUENCE AND ENCODED IC3 AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
AGCGGCTTTA CCTACCGAGA AGAAGAGAGA AAAAAGGTTC GAGTGCGTTC CAGAGTGAGG      60

AGTGAGAAGA G ATG GCT CCG GTG AAG CTG TAC GGC GCG ACC CTG TCG TGG      110
             Met Ala Pro Val Lys Leu Tyr Gly Ala Thr Leu Ser Trp

AAC GTC ACC AGG TGC GTG GCG GCG CTG GAG GAG GCC GGC GTC CAG TAC      158
Asn Val Thr Arg Cys Val Ala Ala Leu Glu Glu Ala Gly Val Gln Tyr

GAG ATC GTA CCC ATC AAC TTC GGC ACC GGC GAG CAC AAG AGC CCC GAC      206
Glu Ile Val Pro Ile Asn Phe Gly Thr Gly Glu His Lys Ser Pro Asp

CAC CTC GCC AGG AAC CCC TTC GGC CAG GTG CCA GCT TTG CAG GAT GGT      254
His Leu Ala Arg Asn Pro Phe Gly Gln Val Pro Ala Leu Gln Asp Gly

GAC TTA TAC GTC TTC GAA TCA CGT GCT ATT TGC AAG TAC GCG TGC CGC      302
Asp Leu Tyr Val Phe Glu Ser Arg Ala Ile Cys Lys Tyr Ala Cys Arg

AAG AAC AAG CCA GAG CTG TTG AAG GAG GGC GAC ATC AAG GAG TCA GCA      350
Lys Asn Lys Pro Glu Leu Leu Lys Glu Gly Asp Ile Lys Glu Ser Ala

ATG GTG GAT GTG TGG CTC GAG GTG GAG GCC CAT CAG TAC ACT GCC GCT      398
Met Val Asp Val Trp Leu Glu Val Glu Ala His Gln Tyr Thr Ala Ala

CTG AGC CCC ATT CTC TTC GAG TGC CTT ATC CAT CCA ATG CTT GGG GGA      446
Leu Ser Pro Ile Leu Phe Glu Cys Leu Ile His Pro Met Leu Gly Gly

GCC ACT GAC CAG AAG GTC ATC GAC GAC AAC CTT GTT AAG ATC AAG AAC      494
Ala Thr Asp Gln Lys Val Ile Asp Asp Asn Leu Val Lys Ile Lys Asn

GTG CTG GCG GTG TAC GAG GCG CAC CTG AGC AAG TCC AAG TAC CTG GCT      542
Val Leu Ala Val Tyr Glu Ala His Leu Ser Lys Ser Lys Tyr Leu Ala

GGA GAC TTC CTC AGT CTT GCG GAC CTT AAC CAT GTG TCT GTC ACC CTG      590
Gly Asp Phe Leu Ser Leu Ala Asp Leu Asn His Val Ser Val Thr Leu

TGC TTG GCG GCT ACA CCC TAT GCG TCT CTG TTC GAC GCG TAC CCG CAT      638
Cys Leu Ala Ala Thr Pro Tyr Ala Ser Leu Phe Asp Ala Tyr Pro His

GTG AAG GCC TGG TGG ACT GAC CTG CTG GCG AGG CCG TCC GTC CAG AAG      686
Val Lys Ala Trp Trp Thr Asp Leu Leu Ala Arg Pro Ser Val Gln Lys

GTC GCA GCG CTG ATG AAG CCA TGATCTTAAT TGCTGGTGCT CGTTCGTCGC         737
Val Ala Ala Leu Met Lys Pro

GAAATAAGCC GAGGTGTGTG CCCCCCGATG TGTGCCTGTA CGAGTGTGTG TTCTTGTGAT   797

GTCTCCTCGT GTTGAATGTT CAGGCTTGTG CTTGCGATCC TGTCTCATCT TTTACTGAAA   857

TGAGCGTTCC TATGCTCTGG TTTAATAATA AATTGTGCCT AGATATTATC TCAAAAAAAA   917
```

AAAAAAAAAA 927

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Met Ala Pro Val Lys Leu Tyr Gly Ala Thr Leu Ser Trp Asn Val Thr
 1               5                  10                  15

Arg Cys Val Ala Ala Leu Glu Glu Ala Gly Val Gln Tyr Glu Ile Val
            20                  25                  30

Pro Ile Asn Phe Gly Thr Gly Glu His Lys Ser Pro Asp His Leu Ala
        35                  40                  45

Arg Asn Pro Phe Gly Gln Val Pro Ala Leu Gln Asp Gly Asp Leu Tyr
    50                  55                  60

Val Phe Glu Ser Arg Ala Ile Cys Lys Tyr Ala Cys Arg Lys Asn Lys
65                  70                  75                  80

Pro Glu Leu Leu Lys Glu Gly Asp Ile Lys Glu Ser Ala Met Val Asp
                85                  90                  95

Val Trp Leu Glu Val Glu Ala His Gln Tyr Thr Ala Ala Leu Ser Pro
            100                 105                 110

Ile Leu Phe Glu Cys Leu Ile His Pro Met Leu Gly Gly Ala Thr Asp
        115                 120                 125

Gln Lys Val Ile Asp Asp Asn Leu Val Lys Ile Lys Asn Val Leu Ala
    130                 135                 140

Val Tyr Glu Ala His Leu Ser Lys Ser Lys Tyr Leu Ala Gly Asp Phe
145                 150                 155                 160

Leu Ser Leu Ala Asp Leu Asn His Val Ser Val Thr Leu Cys Leu Ala
                165                 170                 175

Ala Thr Pro Tyr Ala Ser Leu Phe Asp Ala Tyr Pro His Val Lys Ala
            180                 185                 190

Trp Trp Thr Asp Leu Leu Ala Arg Pro Ser Val Gln Lys Val Ala Ala
        195                 200                 205

Leu Met Lys Pro
    210
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 866 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:45..683

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..866
(D) OTHER INFORMATION:/note= "WIC5 SEQQUENCE AND ENCODED IC5 AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
GAAGCAGGCA ACAGGCGAGC AGGAAGGAAG CAAGAGAGGT GGAG ATG GCG CCC ATC      56
                                                 Met Ala Pro Ile

AAG CTG TAC GGG ATG ATG CTG TCG GCC AAC GTG ACC CGC GTG ACC ACG      104
Lys Leu Tyr Gly Met Met Leu Ser Ala Asn Val Thr Arg Val Thr Thr

CTG CTC AAC GAG CTC GGC CTC GAG TTC GAC TTC GTC GAC GTC GAC CTC      152
Leu Leu Asn Glu Leu Gly Leu Glu Phe Asp Phe Val Asp Val Asp Leu

CGC ACC GGC GCC CAC AAG CAC CCC GAC TTC CTC AAG CTC AAC CCT TTC      200
Arg Thr Gly Ala His Lys His Pro Asp Phe Leu Lys Leu Asn Pro Phe

GGC CAG ATC CCC GCG CTG CAG GAC GGA GAC GAA GTT GTC TTC GAG TCG      248
Gly Gln Ile Pro Ala Leu Gln Asp Gly Asp Glu Val Val Phe Glu Ser

CGC GCC ATC AAC CGG TAC ATC GCG ACC AAG TAC GGG GCG TCC CTG CTG      296
Arg Ala Ile Asn Arg Tyr Ile Ala Thr Lys Tyr Gly Ala Ser Leu Leu

CCG ACG CCG TCG GCC AAG CTG GAG GCG TGG CTG GAG GTG GAG TCG CAC      344
Pro Thr Pro Ser Ala Lys Leu Glu Ala Trp Leu Glu Val Glu Ser His

CAC TTC TAC CCG CCG GCG CGG ACG CTG GTG TAC GAG CTG GTC ATC AAG      392
His Phe Tyr Pro Pro Ala Arg Thr Leu Val Tyr Glu Leu Val Ile Lys

CCC ATG CTG GGC GCC CCC ACC GAC GCC GCC GAG GTG GAC AAG AAC GCC      440
Pro Met Leu Gly Ala Pro Thr Asp Ala Ala Glu Val Asp Lys Asn Ala

GCC GAC CTC GCC AAG CTG CTC GAC GTC TAC GAG GCC CAC CTC GCC GCC      488
Ala Asp Leu Ala Lys Leu Leu Asp Val Tyr Glu Ala His Leu Ala Ala
```

GGG AAC AAG TAC CTG GCC GGC GAC GCC TTC CCG CTC GCC GAC GCC AAC 536
Gly Asn Lys Tyr Leu Ala Gly Asp Ala Phe Pro Leu Ala Asp Ala Asn

CAC ATG TCC TAC CTC TTC ATG CTC ACC AAG AGC CCC AAG GCG GAC CTG 584
His Met Ser Tyr Leu Phe Met Leu Thr Lys Ser Pro Lys Ala Asp Leu

GTG GCC TCC CGC CCG CAC GTC AAG GCC TGG TGG GAG GAG ATC TCC GCC 632
Val Ala Ser Arg Pro His Val Lys Ala Trp Trp Glu Glu Ile Ser Ala

CGC CCC GCC TGG GCC AAG ACC GTC GCC TCC ATC CCC CTC CCG CCC GCC 680
Arg Pro Ala Trp Ala Lys Thr Val Ala Ser Ile Pro Leu Pro Pro Ala

GTC TGAGGTTGCT TGTTTGGCTG CGGCGAGAAC GGAATAAAAT CGCGATGATG 733
Val

GAATAAACAA CTTTTTAGAG AGGAAGCTTG GAATTCTTGG TGTTGCTGCT GTTGAATGTT 793

GAATCTTGGT GTTGAATGTT TACGGCACAT CTAATTTATC CAGTTTTTTT GGCGTGAAAA 853

AAAAAAAAAA AAA 866

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 213 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Ala Pro Ile Lys Leu Tyr Gly Met Met Leu Ser Ala Asn Val Thr
1 5 10 15

Arg Val Thr Thr Leu Leu Asn Glu Leu Gly Leu Glu Phe Asp Phe Val
20 25 30

Asp Val Asp Leu Arg Thr Gly Ala His Lys His Pro Asp Phe Leu Lys
35 40 45

Leu Asn Pro Phe Gly Gln Ile Pro Ala Leu Gln Asp Gly Asp Glu Val
50 55 60

Val Phe Glu Ser Arg Ala Ile Asn Arg Tyr Ile Ala Thr Lys Tyr Gly
65 70 75 80

Ala Ser Leu Leu Pro Thr Pro Ser Ala Lys Leu Glu Ala Trp Leu Glu
85 90 95

Val Glu Ser His His Phe Tyr Pro Pro Ala Arg Thr Leu Val Tyr Glu
100 105 110

Leu Val Ile Lys Pro Met Leu Gly Ala Pro Thr Asp Ala Ala Glu Val

```
        115                 120                 125

Asp Lys Asn Ala Ala Asp Leu Ala Lys Leu Leu Asp Val Tyr Glu Ala
    130                 135                 140

His Leu Ala Ala Gly Asn Lys Tyr Leu Ala Gly Asp Ala Phe Pro Leu
145                 150                 155                 160

Ala Asp Ala Asn His Met Ser Tyr Leu Phe Met Leu Thr Lys Ser Pro
                165                 170                 175

Lys Ala Asp Leu Val Ala Ser Arg Pro His Val Lys Ala Trp Trp Glu
            180                 185                 190

Glu Ile Ser Ala Arg Pro Ala Trp Ala Lys Thr Val Ala Ser Ile Pro
        195                 200                 205

Leu Pro Pro Ala Val
    210
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 897 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:15..668

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..897
(D) OTHER INFORMATION:/note= "WIC4 SEQUENCE AND ENCODED IC4 AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
AACCAAGGGA AACA ATG GCG CCG GTG AAG GTG TTC GGG CCG GCG ATG TCG        50
                Met Ala Pro Val Lys Val Phe Gly Pro Ala Met Ser

ACC AAC GTG GCC CGG GTG CTG GTG TGC CTG GAG GAG GTC GGC GCC GAG        98
Thr Asn Val Ala Arg Val Leu Val Cys Leu Glu Glu Val Gly Ala Glu

TAC GAG GTG GTC GAC ATC GAT TTC AAG GCC ATG GAG CAC AAG AGC CCC       146
Tyr Glu Val Val Asp Ile Asp Phe Lys Ala Met Glu His Lys Ser Pro

GAG CAT CTC GTC AGA AAC CCG TTC GGC CAA ATC CCT GCC TTC CAG GAT       194
Glu His Leu Val Arg Asn Pro Phe Gly Gln Ile Pro Ala Phe Gln Asp
```

```
GGG GAT CTG CTT CTC TTC GAG TCA CGC GCA ATT GCG AGG TAC GTG CTC      242
Gly Asp Leu Leu Leu Phe Glu Ser Arg Ala Ile Ala Arg Tyr Val Leu

CGC AAG TAC AAG AAG AAC GAA GTG GAC CTG CTG AGG GAA GGC GAC CTC      290
Arg Lys Tyr Lys Lys Asn Glu Val Asp Leu Leu Arg Glu Gly Asp Leu

AAG GAG GCG GCG ATG GTG GAC GTA TGG ACG GAG GTG GAC GCG CAC ACC      338
Lys Glu Ala Ala Met Val Asp Val Trp Thr Glu Val Asp Ala His Thr

TAC AAC CCG GCC ATC TCG CCG ATC GTG TAC GAG TGC TCA TCA ACC GCT      386
Tyr Asn Pro Ala Ile Ser Pro Ile Val Tyr Glu Cys Ser Ser Thr Ala

CAT GCG CGG CTG CCG ACC AAC CAA ACG GTG GTG GAC GAG AGC CTG GAG      434
His Ala Arg Leu Pro Thr Asn Gln Thr Val Val Asp Glu Ser Leu Glu

AAG CTC AAG AAC GTG CTG GAG GTC TAC GAG GCG CGC CTG TCC AAG CAC      482
Lys Leu Lys Asn Val Leu Glu Val Tyr Glu Ala Arg Leu Ser Lys His

GAC TAC CTC GCC GGG GAC TTC GTC AGC TTC GCG GAC CTC AAC CAC TTC      530
Asp Tyr Leu Ala Gly Asp Phe Val Ser Phe Ala Asp Leu Asn His Phe

CCC TAC ACC TTC TAC TTC ATG GCC ACG CCG CAC GCG GCC CTC TTC GAC      578
Pro Tyr Thr Phe Tyr Phe Met Ala Thr Pro His Ala Ala Leu Phe Asp

TCG TAC CCG CAC GTC AAG GCC TGG TGG GAG AGG ATC ATG GCG AGG CCG      626
Ser Tyr Pro His Val Lys Ala Trp Trp Glu Arg Ile Met Ala Arg Pro

GCC GTG AAG AAG CTC GCC GCG CAG ATG GTT CCC AAG AAG CCG              668
Ala Val Lys Lys Leu Ala Ala Gln Met Val Pro Lys Lys Pro

TGATTTGCTA GGCGGGATCT CGCATCGTGG GATCCGATTC CGATCACTGA TCTGTGTGGC    728

GTTTTCTTTT CTTGTTGGTG TCGCGAATAA GGCAAATGAG CTCGTGTGTG TGTGGCTGGA    788

ATTGCACCAG CGTGCAGTTT TTGCGCTTTG CGTGTGTGTG GTCGTGAAAA CTCTTGAGAT    848

GGAACAATGT CTTCGTAATG CTTTCACATT TTAAAAAAAA AAAAAAAAA                897
```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
Met Ala Pro Val Lys Val Phe Gly Pro Ala Met Ser Thr Asn Val Ala
 1               5                  10                  15

Arg Val Leu Val Cys Leu Glu Glu Val Gly Ala Glu Tyr Glu Val Val
            20                  25                  30
```

```
Asp Ile Asp Phe Lys Ala Met Glu His Lys Ser Pro Glu His Leu Val
         35                  40                  45

Arg Asn Pro Phe Gly Gln Ile Pro Ala Phe Gln Asp Gly Asp Leu Leu
     50                  55                  60

Leu Phe Glu Ser Arg Ala Ile Ala Arg Tyr Val Leu Arg Lys Tyr Lys
 65                  70                  75                  80

Lys Asn Glu Val Asp Leu Leu Arg Glu Gly Asp Leu Lys Glu Ala Ala
                 85                  90                  95

Met Val Asp Val Trp Thr Glu Val Asp Ala His Thr Tyr Asn Pro Ala
            100                 105                 110

Ile Ser Pro Ile Val Tyr Glu Cys Ser Ser Thr Ala His Ala Arg Leu
        115                 120                 125

Pro Thr Asn Gln Thr Val Val Asp Glu Ser Leu Glu Lys Leu Lys Asn
    130                 135                 140

Val Leu Glu Val Tyr Glu Ala Arg Leu Ser Lys His Asp Tyr Leu Ala
145                 150                 155                 160

Gly Asp Phe Val Ser Phe Ala Asp Leu Asn His Phe Pro Tyr Thr Phe
                165                 170                 175

Tyr Phe Met Ala Thr Pro His Ala Ala Leu Phe Asp Ser Tyr Pro His
            180                 185                 190

Val Lys Ala Trp Trp Glu Arg Ile Met Ala Arg Pro Ala Val Lys Lys
        195                 200                 205

Leu Ala Ala Gln Met Val Pro Lys Lys Pro
    210                 215
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 721 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:21..686

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..721

(D) OTHER INFORMATION:/note= "TA 27 SEQUENCE AND ENCODED AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
TTCGGCACGA GGAAGAAGGG ATG GAG CCT ATG AAG GTG TAC GGC TGG GCG          50
                      Met Glu Pro Met Lys Val Tyr Gly Trp Ala

GTG TCG CCA TGG ATG GCG CGG GTC CTC GTC TCC CTG GAG GAG GCC GGC        98
Val Ser Pro Trp Met Ala Arg Val Leu Val Ser Leu Glu Glu Ala Gly

GCC GAC TAC GAG CTC GTG CCC ATG AGC CGC AAC GGC GGC GAC CAC CGG       146
Ala Asp Tyr Glu Leu Val Pro Met Ser Arg Asn Gly Gly Asp His Arg

CGG CCG GAG CAC CTC GCC AGA AAC CCC TTC GGT GAG ATC CCG GTG CTC       194
Arg Pro Glu His Leu Ala Arg Asn Pro Phe Gly Glu Ile Pro Val Leu

GAA TAC GGC GGT CTG ACG CTT TAC CAA TCC CGC GCC ATT GCA AGG CAT       242
Glu Tyr Gly Gly Leu Thr Leu Tyr Gln Ser Arg Ala Ile Ala Arg His

ATT CTC CGC AAA CAC AAG CCC GGG CTT CTA GGA GCA GGC AGC CTC GAG       290
Ile Leu Arg Lys His Lys Pro Gly Leu Leu Gly Ala Gly Ser Leu Glu

GAG TCG GCG ATG GTG GAT GTA TGG GTC GAC GTG GAT GCC CAC CAC CTG       338
Glu Ser Ala Met Val Asp Val Trp Val Asp Val Asp Ala His His Leu

GAG CCC GTA CTC AAG CCC ATC GTG TGG AAC TGC ATC ATC AAC CCG TTC       386
Glu Pro Val Leu Lys Pro Ile Val Trp Asn Cys Ile Ile Asn Pro Phe

GTC GGG AGG GAC GTC GAC CAG GGC CTC GTC GAT GAG AGC GTC GAG AAG       434
Val Gly Arg Asp Val Asp Gln Gly Leu Val Asp Glu Ser Val Glu Lys

CTC AAG AAG CTG CTG GAG GTG TAC GAG GCA AGA CTG TCA AGC AAC AAG       482
Leu Lys Lys Leu Leu Glu Val Tyr Glu Ala Arg Leu Ser Ser Asn Lys

TAC TTG GCC GGG GAT TTC GTC AGC TTC GCC GAC CTC ACC CAT TTC TCC       530
Tyr Leu Ala Gly Asp Phe Val Ser Phe Ala Asp Leu Thr His Phe Ser

TTC ATG CGC TAC TTC ATG GCG ACG GAG CAT GCG GTT GTG CTC GAT GCG       578
Phe Met Arg Tyr Phe Met Ala Thr Glu His Ala Val Val Leu Asp Ala

TAT CCG CAT GTG AAG GCA TGG TGG AAG GCG CTG CTG GCA AGG CCA TCG       626
Tyr Pro His Val Lys Ala Trp Trp Lys Ala Leu Leu Ala Arg Pro Ser

GTC AAG AAG GTG ATA GCT GGC ATG CCT CCG GAT TTT GGA TTC GGG AGC       674
Val Lys Lys Val Ile Ala Gly Met Pro Pro Asp Phe Gly Phe Gly Ser

GGG AGA ATA CCA TGATAAAGCA TGCTTGTTTG TCTATGATGC TCTGA               721
Gly Arg Ile Pro
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
Met Glu Pro Met Lys Val Tyr Gly Trp Ala Val Ser Pro Trp Met Ala
 1               5                  10                  15

Arg Val Leu Val Ser Leu Glu Glu Ala Gly Ala Asp Tyr Glu Leu Val
            20                  25                  30

Pro Met Ser Arg Asn Gly Gly Asp His Arg Arg Pro Glu His Leu Ala
        35                  40                  45

Arg Asn Pro Phe Gly Glu Ile Pro Val Leu Glu Tyr Gly Gly Leu Thr
    50                  55                  60

Leu Tyr Gln Ser Arg Ala Ile Ala Arg His Ile Leu Arg Lys His Lys
65                  70                  75                  80

Pro Gly Leu Leu Gly Ala Gly Ser Leu Glu Glu Ser Ala Met Val Asp
                85                  90                  95

Val Trp Val Asp Val Asp Ala His His Leu Glu Pro Val Leu Lys Pro
            100                 105                 110

Ile Val Trp Asn Cys Ile Ile Asn Pro Phe Val Gly Arg Asp Val Asp
        115                 120                 125

Gln Gly Leu Val Asp Glu Ser Val Glu Lys Leu Lys Lys Leu Leu Glu
    130                 135                 140

Val Tyr Glu Ala Arg Leu Ser Ser Asn Lys Tyr Leu Ala Gly Asp Phe
145                 150                 155                 160

Val Ser Phe Ala Asp Leu Thr His Phe Ser Phe Met Arg Tyr Phe Met
                165                 170                 175

Ala Thr Glu His Ala Val Val Leu Asp Ala Tyr Pro His Val Lys Ala
            180                 185                 190

Trp Trp Lys Ala Leu Leu Ala Arg Pro Ser Val Lys Lys Val Ile Ala
        195                 200                 205

Gly Met Pro Pro Asp Phe Gly Phe Gly Ser Gly Arg Ile Pro
    210                 215                 220
```

CLAIMS

1. A polynucleotide encoding a glutathione transferase (GST) subunit, which polynucleotide comprises a coding sequence capable of hybridising selectively to the coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 or to the complement of one of those sequences.

2. A polynucleotide of claim 1 which is a DNA sequence.

3. A polynucleotide according to claim 1 or 2 wherein the coding sequence encodes the amino acid sequence of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14.

4. A polynucleotide according to any one of the preceding claims which comprises the coding sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13 or a fragment thereof.

5. A polypeptide which is a GST subunit and comprises the amino acid sequence of SEQ ID No. 2, 4, 6, 8, 10, 12 or 14 or a sequence substantially homologous thereto, or a fragment of either said sequence.

6. A polypeptide according to claim 5 encoded by the coding sequence of a polynucleotide according to any one of claims 1 to 4.

7. A dimeric protein comprising two GST subunits, wherein at least one subunit is a polypeptide according to claim 5 or 6.

8. A chimeric gene comprising a polynucleotide according to any one of claims 1 to 4 operably linked to regulatory sequences that allow expression of the coding sequence in a host cell.

9. A chimeric gene according to claim 7 wherein the regulatory sequences allow expression of the coding sequence in a plant cell.

10. A vector comprising a polynucleotide according to any one of claims 1 to 4 or a chimeric gene according to claim 8 or 9.

11. A vector according to claim 10 which is an expression vector.

12. A cell transformed or transfected with a vector according to claim 10 or 11.

13. A cell according to claim 12 which is a prokaryotic cell or a plant cell.

14. A cell having, integrated into its genome, a chimeric gene according to claim 8 or 9.

15. A cell according to claim 14 which is a plant cell, wherein the chimeric gene is a chimeric gene according to claim 9.

16. A cell according to any one of claims 12 to 15 further comprising one or more further polynucleotide sequences coding for a GST subunit, operably linked to regulatory elements that allow expression of the subunit in the cell.

17. A process for the production of a polypeptide according to claim 5 or 6, which process comprises:

(a) cultivating a cell according to any one of claims 12 to 15 under conditions that allow the expression of the polypeptide; and

(b) recovering the expressed polypeptide.

18. A process for the production of a dimeric protein according to claim 7, which process comprises:

(a) cultivating a cell according to any one of claims 12 to 16 under conditions that allow:

(i) the expression of the polypeptide according to claim 5 or 6 and, if a further polynucleotide sequence as defined in claim 16 is present, optionally the expression of a further GST subunit encoded by a further polynucleotide, and

(ii)the association of the GST subunit polypeptide according to claim 5 or 6 with another GST subunit polypeptide to form a dimeric protein according to claim 7; and

(b) recovering the dimeric protein so formed.

19. A process according to claim 17 or 18 wherein the cell is a prokaryotic cell or a plant cell.

20. A method of obtaining a transgenic plant cell comprising:

(a) transforming a plant cell with an expression vector according to claim 11 to give a transgenic plant cell,

and optionally,

(a') transforming the cell with one or more further polynucleotide sequences coding for a GST subunit, operably linked to regulatory elements that allow expression of the subunit in the cell.

21. A method of obtaining a first-generation transgenic plant comprising:

(b) regenerating a transgenic plant cell transformed with a vector according to claim 11 to give a transgenic plant.

22. A method of obtaining a transgenic plant seed comprising:

(c) obtaining a transgenic seed from a transgenic plant obtainable by step (b) of claim 21.

23. A method of obtaining a transgenic progeny plant comprising obtaining a second-generation transgenic progeny plant from a first-generation transgenic plant obtainable by a method according to claim 21, and optionally obtaining transgenic plants of one or more further generations from the second-generation progeny plant thus obtained.

24. A method according to claim 23 comprising:

(c) obtaining a transgenic seed from a first-generation transgenic plant obtainable by the method according to claim 21, then obtaining a second-generation transgenic progeny plant from the transgenic seed;

and/or

(d) propagating clonally a first-generation transgenic plant obtainable by the method according to claim 21 to give a second-generation progeny plant;

and/or

(e) crossing a first-generation transgenic plant obtainable by a method according to claim 21 with another plant to give a second-generation progeny plant;

and optionally;

(f) obtaining transgenic progeny plants of one or more further generations from the second-generation progeny plant thus obtained.

25. A transgenic plant cell, first-generation plant, plant seed or progeny plant obtainable by a method according to any one of claims 20 to 24.

26. A transgenic plant or plant seed comprising plant cells according to claim 13 or 15.

27. A transgenic plant cell callus comprising plant cells according to claim 13 or 15, or obtainable from a transgenic plant cell, first-generation plant, plant seed or progeny plant according to claim 25.

28. Use of a polynucleotide according to any one of claims 1 to 4 as a selectable marker for detecting transformation of a plant cell.

29. A nucleic acid construct comprising:

(a) a polynucleotide according to any one of claims 1 to 4 operably linked to regulatory elements that allow expression of the coding sequence in a plant cell; and

(b) a site into which a further polynucleotide comprising a coding sequence can be inserted.

30. A nucleic acid construct according to claim 29 wherein site (b) is bounded by regulatory elements that allow expression of a coding sequence inserted at the site in a plant cell.

31. A vector comprising a construct according to claim 29.

32. A method of transforming a plant cell or of obtaining a plant cell culture or transgenic plant comprising:

(a) providing an untransformed plant cell which is susceptible to a herbicide whose herbicidal activity is reduced by a dimeric protein according to claim 7;

(b) transforming the plant cell with a vector according to claim 29 or 30;

(c) cultivating the transformed cell under conditions that allow the expression of the polynucleotide (a) in the construct according to claim 29 or 30; and/or

(c') regenerating the cell to give a cell culture or plant such that the polynucleotide (a) in the construct according to claim 29 or 30 is expressed; and

(d) contacting the cell, cell culture or plant with the herbicide whose herbicidal activity is reduced by the dimeric protein according to claim 7, and to which the untransformed plant cell was susceptible; and

(e) selecting cells, cell cultures or plants that are less susceptible to the herbicide than are corresponding untransformed cells, cell cultures or plants.

33. Use of a dimeric protein according to claim 7 in a method of identifying compounds capable of metabolism by a GST.

34. A method of identifying compounds capable of being metabolised by a glutathione transferase comprising:

(a) contacting a candidate compound suspected of being capable of being metabolised by glutathione transferase with glutathione (GSH) in the presence of a dimeric protein according to claim 7; and

(b) determining whether or not metabolism of the candidate compound takes place.

35. A method according to claim 34 wherein metabolism of the compound is detected by determining whether or not it is conjugated to glutathione by the dimeric protein according to claim 7.

36. A kit for detecting compounds capable of being metabolised by a GST comprising:

(a) reduced glutathione or homoglutathione; and

(b) a dimeric protein according to claim 7.

37. An antibody which specifically recognises a polypeptide according to claim 5 or 6 or a dimeric protein according to claim 7.

38. A nucleic acid probe which selectively hybridises to the sequence of SEQ ID No. 1, 3, 5, 7, 9, 11 or 13.

39. A method of identifying compounds that induce GST expression in graminaceous plants comprising:

(a) contacting a graminaceous plant, or a cell or cell culture thereof, with a candidate compound suspected of being capable of inducing GST expression; and

(b) determining the level of GST expression in the

plant, cell or cell culture.

40. A method according to claim 39 wherein, in step (b), the level of expression is determined by: (i) determining the level of GST protein present by using an antibody according to claim 35; or (ii) determining the level of GST mRNA present using a probe according to claim 37.

41. A kit for identifying compounds that induce GST expression in plants by a method as defined in claim 37 or 38, which kit comprises an antibody as defined in claim 36.

42. A method of determining the GST level in a sample of seed or flour comprising:

(i) determining the level of GST protein present by using an antibody according to claim 35; or

(ii) determining the level of GST mRNA present using a probe according to claim 37.

43. A method of controlling the growth of weeds at a locus where a transgenic plant according to any one of claims 25 to 27 is being cultivated, which method comprises applying to the locus a herbicide whose herbicidal properties are reduced by a dimeric protein according to claim 7.

44. A compound identified by a method according to any one of claims 34, 35, 39 or 40.

45. A polynucleotide according to claim 1 substantially as hereinbefore described with reference to any one of the preceding Examples.

46. A polypeptide according to claim 5 substantially as hereinbefore described with reference to any one of the preceding Examples.

47. A dimeric protein according to claim 7 substantially as hereinbefore described with reference to any one of the preceding Examples.

48. A chimeric gene according to claim 8 substantially as hereinbefore described with reference to any one of the preceding Examples.

49. A vector according to claim 10 substantially as hereinbefore described with reference to any one of the preceding Examples.

50. A cell according to claim 12 substantially as hereinbefore described with reference to any one of the preceding Examples.

51. A process according to claim 17 or 18 substantially as hereinbefore described with reference to any one of the preceding Examples.

52. A method according to claims 20, 21, 22 or 23 substantially as hereinbefore described with reference to any one of the preceding Examples.

53. A transgenic plant cell, first-generation plant, plant seed or progeny plant, plant or plant seed, or plant cell callus according to any one of claims 25 to 27 substantially as hereinbefore described with reference to any one of the preceding Examples.

54. Use according to claim 28 substantially as hereinbefore described with reference to any one of the preceding Examples.

55. A nucleic acid construct according to claim 29 substantially as hereinbefore described with reference to any one of the preceding Examples.

56. A vector according to claim 31 substantially as hereinbefore described with reference to any one of the preceding Examples.

57. A method according to claim 32 substantially as hereinbefore described with reference to any one of the preceding Examples.

58. Use according to claim 33 substantially as hereinbefore described with reference to any one of the preceding Examples.

59. A method according to claim 34 substantially as hereinbefore described with reference to any one of the preceding Examples.

60. An antibody according to claim 37 substantially as hereinbefore described with reference to any one of the preceding Examples.

61. A nucleic acid probe according to claim 38 substantially as hereinbefore described with reference to any one of the preceding Examples.

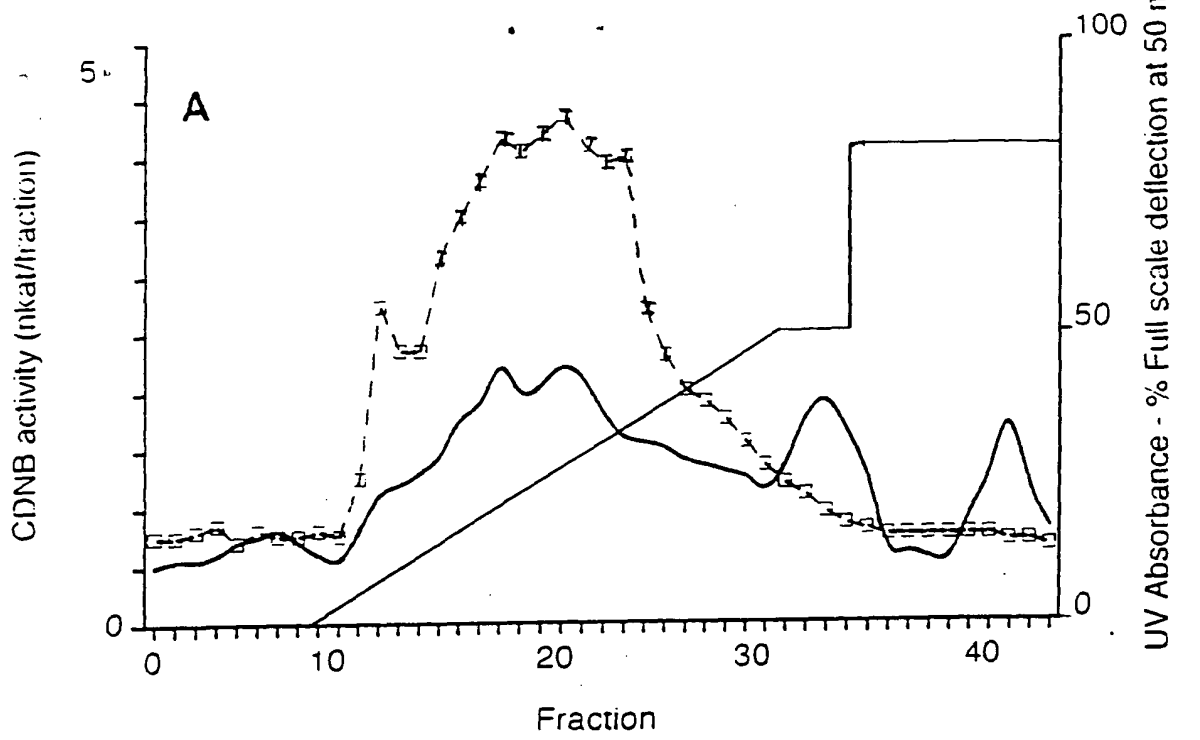
62. A method according to claims 39, 42 or 43 substantially as hereinbefore described with reference to any one of the preceding Examples or 42.

63. A compound according to claim 44 substantially as hereinbefore described with reference to any one of the preceding Examples.

ABSTRACT

NEW PLANT GENES

This invention relates to glutathione transferase (GST) subunits to nucleic acid sequences encoding glutathione transferase subunits, and to uses of these glutathione transferases and coding sequences, especially in the field of plant biotechnology.


A
CDNB activity (nkat/fraction)
UV Absorbance - % Full scale deflection at 50 mV
Fraction
0
10
20
30
40
50
100

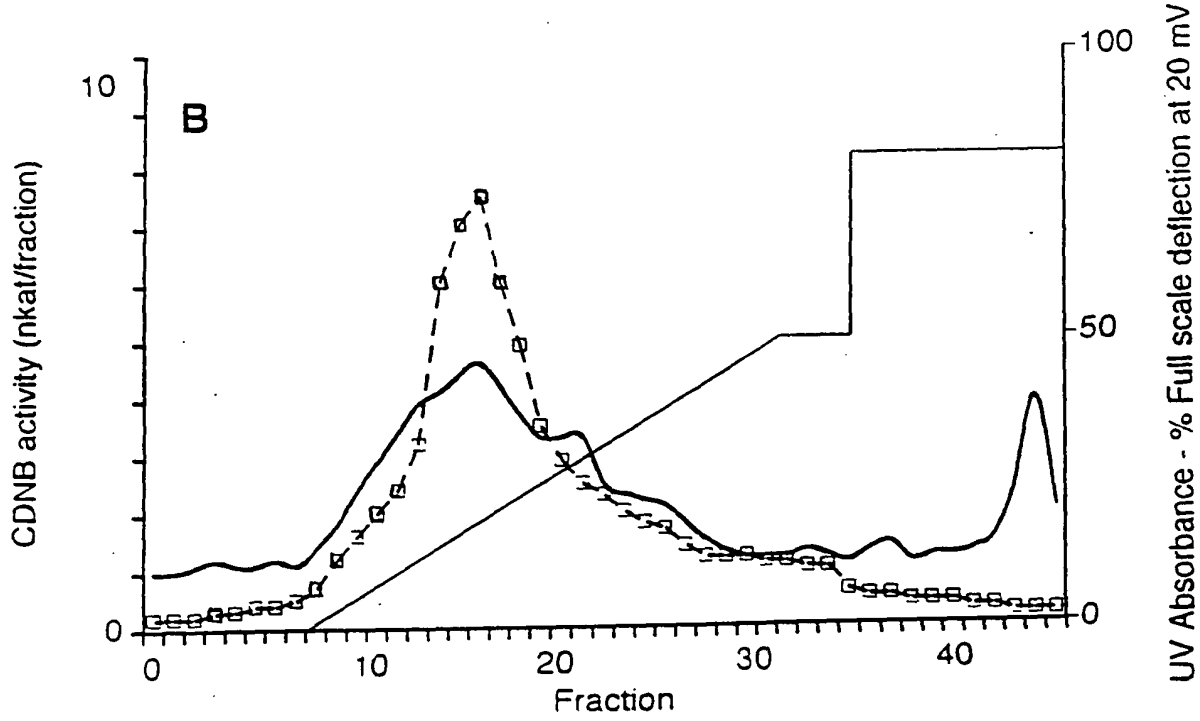

B
CDNB activity (nkat/fraction)
UV Absorbance - % Full scale deflection at 20 mV
Fraction
0
10
20
30
40
50
100

Figure 1

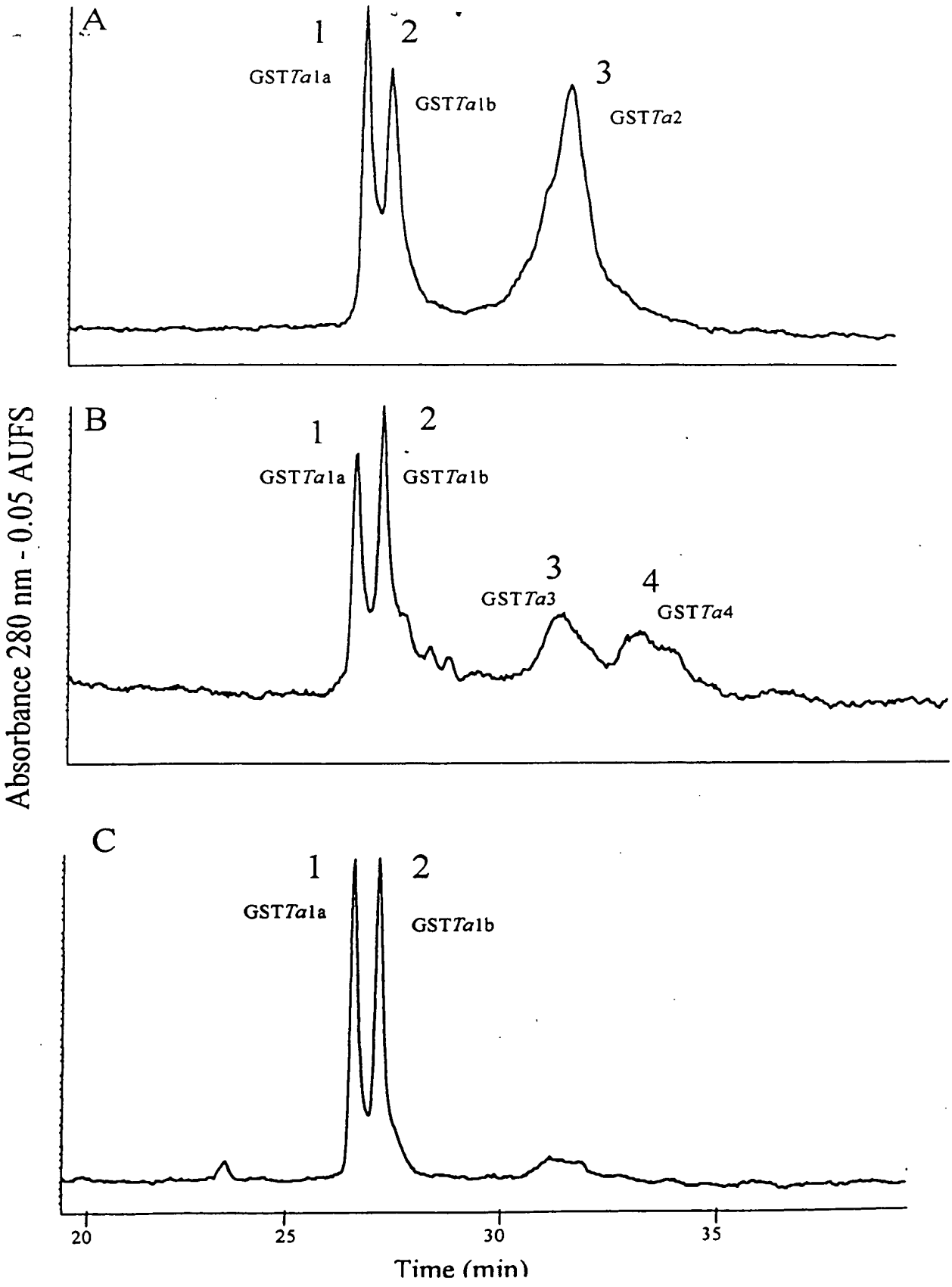

A
1
GSTTa1a
2
GSTTa1b
3
GSTTa2
B
1
GSTTa1a
2
GSTTa1b
3
GSTTa3
4
GSTTa4
Absorbance 280 nm - 0.05 AUFS
C
1
GSTTa1a
2
GSTTa1b
20
25
30
35
Time (min)

Figure 2

PCT NO : [redacted]98 / 02802

Form 23/77 : 15/01/98

Agent : J A KEMP & CO